

Isoform 1:

```

1  TTGCTCACTG CTCACCCACC TGCTGCTGCC ATGAGGCACC TTGGGGCCTT
51 CCTCTTCCTT CTGGGGGTCC TGGGGGCCCT CACTGAGATG TGTGAAATAC
101 CAGAGATGGA CAGCCATCTG GTAGAGAAGT TGGGCCAGCA CCTCTTACCT
151 TGGATGGACC GGCTTTCCCT GGAGCACTTG AACCCAGCA TCTATGTGGG
201 CCTACGCCTC TCCAGTCTGC AGGCTGGGAC CAAGGAAGAC CTCTACCTGC
251 ACAGCCTCAA GCTTGGTTAC CAGCAGTGCC TCCTAGGGTC TGCCTTCAGC
301 GAGGATGACG GTGACTGCCA GGGCAAGCCT TCCATGGGCC AGCTGGCCCT
351 CTACCTGCTC GCTCTCAGAG CCAACTGTGA GTTTGTGAGG GGCCACAAGG
401 GGGACAGGCT GGTCTCACAG CTCAAATGGT TCCTGGAGGA TGAGAAGAGA
451 GCCATTGACA CAGCAGCCAT GGCAGGCTTG GCATTACCT GTCTGAAGCG
501 CTCAAACCTC AACCTGGTC GGAGACAACG GATCACCATG GCCATCAGAA
551 CAGTGCGAGA GGAGATCTTG AAGGCCAGA CCCCCAGGG CCACTTTGGG
601 AATGTCTACA GCACCCCAT GGCATTACAG TTCCTCATGA CTTCCCCCAT
651 GCGTGGGGCA GAACTGGGAA CAGCATGTCT CAAGGCGAGG GTTGCTTTGC
701 TGGCCAGTCT GCAGGATGGA GCCTTCCAGA ATGCTCTCAT GATTCCCG
751 CTGCTGCCCG TTCTGAACCA CAAGACCTAC ATTGATCTGA TCTTCCAGA
801 CTGTCTGGCA CCACGAGTCA TGTTGGAACC AGCTGCTGAG ACCATTCCCTC
851 AGACCAAGA GATCATCAGT GTCACGCTGC AGGTGCTTAG TCTCTGCCC
901 CCGTACAGAC AGTCCATCTC TGTTCTGGCC GGTCCACCG TGGAAGATGT
951 CCTGAAGAAG GCCCATGAGT TAGGAGGATT CACATATGAA ACACAGGCC
1001 CCTTGTCAGG CCCCTACTTA ACCTCCGTGA TGGGAAAGC GGCCGGAGAA
1051 AGGGAGTTCT GGCAGCTTCT CCGAGACCCC AACACCCAC TGTGCAAGG
1101 TATTGCTGAC TACAGACCCA AGGATGGAGA AACCATTGAG CTGAGGCTGG
1151 TTAGCTGGTA GCCCTGAGC TCCCTCATCC CAGCAGCCTC GCACACTCCC
1201 TAGGCTTCTA CCCTCCCTCC TGATGTCCCT GGAACAGGAA CTCGCCTGAC
1251 CCTGCTGCCA CCTCCTGTGC ACTTTGAGCA ATGCCCCCTG GGATCACCCC
1301 AGCCACAAGC CCTTCGAGG CCCTATACCA TGGCCACCT TGGAGCAGAG
1351 AGCCAAAGCAT CTTCCCTGGG AAGTCTTTCT GGCCAAGTCT GGCCAGCCTG
1401 GCCCTGCAGG TCTCCCATGA AGGCCACCCC ATGGTCTGAT GGGCATGAAG
1451 CATCTCAGAC TCCTTGGAAC AAAACGGAGT CCGCAGGCCG CAGGTGTTGT
1501 GAAGACCACT CGTTCTGTGG TTGGGGTCCT GCAAGAAGGC CTCCTCAGCC
1551 CGGGGGCTAT GGCCCTGACC CCAGCTCTCC ACTCTGCTGT TAGAGTGGCA
1601 GCTCCGAGCT GGTGTGGCA CAGTAGCTGG GGAGACCTCA GCAGGGCTGC
1651 TCAGTGCCTG CCTCTGACAA AATTAAAGCA TTGATGGCCT GTGAAAAAAA
1701 AAAAAAAAAA AAAAAAAAAA AA

```

(SEQ ID NO:1)

FEATURES:

5'UTR: 1 - 30
Start Codon: 31
Stop Codon: 1159
3'UTR: 1162

Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA 108000024653390 /altid=gi 12742775 /def=ref XP_009922.2 tr...	752	0.0
CRA 108000024636236 /altid=gi 298316 /def=gb AAB25526.1 transc...	732	0.0
CRA 18000004926133 /altid=gi 339205 /def=gb AAA61057.1 (L02648...	732	0.0
CRA 108000024042036 /altid=gi 12654675 /def=gb AAH01176.1 AAH01...	731	0.0
CRA 18000004926130 /altid=gi 4507409 /def=ref NP_000346.1 tran...	727	0.0
CRA 18000004926132 /altid=gi 339203 /def=gb AAA61056.1 (L02647...	725	0.0
CRA 18000005170902 /altid=gi 7657639 /def=ref NP_056564.1 tran...	515	e-145
CRA 18000005218941 /altid=gi 4572454 /def=gb AAD23829.1 AF12128...	501	e-140
CRA 164000136745249 /altid=gi 11968124 /def=ref NP_071979.1 tr...	481	e-134
CRA 18000004926134 /altid=gi 4507407 /def=ref NP_001053.1 tran...	108	2e-22

EST:

gi 10725490 /dataset=dbest /taxon=96...	858	0.0
gi 10947399 /dataset=dbest /taxon=96...	846	0.0
gi 9121897 /dataset=dbest /taxon=9606...	846	0.0
gi 13280819 /dataset=dbest /taxon=96...	846	0.0
gi 13287907 /dataset=dbest /taxon=96...	833	0.0
gi 13286505 /dataset=dbest /taxon=96...	831	0.0
gi 8150776 /dataset=dbest /taxon=960...	815	0.0
gi 5936410 /dataset=dbest /taxon=9606 ...	726	0.0
gi 6888875 /dataset=dbest /taxon=9606...	726	0.0
gi 6888872 /dataset=dbest /taxon=9606...	726	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

gi|10725490|adult adrenal gland
gi|10947399| mammary gland
gi|9121897| retinoblastoma
gi|13280819| adenocarcinoma cell line
gi|13287907| retinoblastoma
gi|13286505| embryonal carcinoma, cell line
gi|8150776| adult uterus
gi|5936410| adult uterus
gi|6888875| adult head_neck
gi|6888872| adult head_neck

Tissue Expression:

Human leukocyte

Isoform 2:

```

1  GGAGGATTAA TCAGTGACAG GAAGCTGCGT CTCTCGGAGC GGTGACCAGC
51  TGTGGTCAGG AGAGCCTCAG CAGGGCCAGC CCCAGGAGTC TTTCCCGATT
101 CTGTGCTCACT GCTCACCCAC CTGCTGCTGC CATGAGGCAC CTTGGGGCCT
151 TCCTCTTCCT TCTGGGGGTC CTGGGGGCCC TCACTGAGAT GTGTGAAATA
201 CCAGAGATGG ACAGCCATCT GGTAGAGAAG TTGGGCCAGC ACCTCTTACC
251 TTGGATGGAC CGGCTTTCCT TGGAGCACTT GAACCCAGC ATCTATGTGG
301 GCCTACGCCT CTCCAGTCTG CAGGCTGGGA CCAAGGAAGA CCTCTACCTG
351 CACAGCCTCA TGCTTGGTTA CCAGCAGTGC CTCCTAGGGT CTGCCTTCAG
401 CGAGGATGAC GGTGACTGCC AGGGCAAGCC TTCCATGGGC CAGCTGGCCC
451 TCTACCTGCT CGCTCTCAGA GCCAACTGGC ATGATCACAA GGGCCACCCC
501 CACACTAGCT ACTACCAGTA TGGCCTGGGC ATTCTGGCCC TGTGTCTCCA
551 CCAGAAGCGG GTCCATGACA GCGTGGTGGG CAAACTTCTG TATGCTGTGG
601 AACCTTTCCA CCAGGGCCAC CATTCTGTGG ACACAGCAGC CATGGCAGGC
651 TTGGCATTCA CCTGTCTGAA GCGCTCAAAC TTCAACCCTG GTCGGAGACA
701 ACGGATCACC ATGGCCATCA GAACAGTGGC AGAGGAGATC TTGAAGGCCC
751 AGACCCCGA GGGCCACTTT GGAATGTCT ACAGCACCCC ATTGGCATTG
801 CAGTTCTCTA TGACTTCCCC CATGCGTGGG GCAGAACTGG GAACAGCATG
851 TCTCAAGGCG AGGGTTGCTT TGCTGGCCAG TCTGCAGGAT GGAGCCTTCC
901 AGAATGCTCT CATGATTTCC CAGCTGCTGC CCGTTCTGAA CCACAAGACC
951 TACATTGATC TGATCTTCCC AGACTGTCTG GCACCACGAG TCATGTTGGA
1001 ACCAGCTGCT GAGACCATTG CTCAGACCCA AGAGATCATC AGTGTCAAGC
1051 TGCAGGTGCT TAGTCTCTTG CCGCCGTACA GACAGTCCAT CTCTGTTCTG
1101 GCCGGGTCCA CCGTGAAGA TGTCTGAAG AAGGCCCATG AGTTAGGAGG
1151 ATTACATAT GAAACACAGG CCTCCTTGTG AGGCCCTAC TTAACCTCCG
1201 TGATGGGGAA AGCGGCCGGA GAAAGGGAGT TCTGGCAGCT TCTCCGAGAC
1251 CCAAACACCC CACTGTTGCA AGGTATTGCT GACTACAGAC CCAAGGATGG
1301 AGAAACCATG GAGCTGAGGC TGGTTAGCTG GTAGCCCTG AGCTCCCTCA
1351 TCCCAGCAGC CTCGCACACT CCCTAGGCTT CTACCCCTCC TCCTGATGTC
1401 CCTGGAACAG GAACTCGCCT GACCCTGCTG CCACCTCCTG TGCACTTTGA
1451 GCAATGCCCC CTGGGATCAC CCCAGCCACA AGCCCTTCGA GGGCCCTATA
1501 CCATGGCCCA CTTGGAGCA GAGAGCCAAG CATCTTCCCT GGGAAAGTCTT
1551 TCTGGCCAAG TCTGGCCAGC CTGGCCCTGC AGGTCTCCCA TGAAGGCCAC
1601 CCCATGGTCT GATGGGCATG AAGCATCTCA GACTCCTTGG CAAAAACGG
1651 AGTCCGAGG CCGCAGGTGT TGTGAAGACC ACTCGTTCCT TGGTTGGGGT
1701 CCTGCAAGAA GGCCTCCTCA GCCCAGGGGC TATGGCCCTG ACCCCAGCTC
1751 TCCACTCTGC TGTTAGAGTG GCAGCTCCGA GCTGGTTGTG GCACAGTAGC
1801 TGGGGAGACC TCAGCAGGGC TGCTCAGTGC CTGCCTCTGA CAAAATTAAA
1851 GCATTGATGG CCTGTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA

```

(SEQ ID NO:2)

FEATURES:

5'UTR: 1 - 131
Start Codon: 132
Stop Codon: 1332
3'UTR: 1335

Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA 108000024636236 /altid=gi 298316 /def=gb AAB25526.1 transc...	793	0.0
CRA 108000024653390 /altid=gi 12742775 /def=ref XP_009922.2 tr...	793	0.0
CRA 18000004926133 /altid=gi 339205 /def=gb AAA61057.1 (L02648...	792	0.0
CRA 108000024042036 /altid=gi 12654675 /def=gb AAH01176.1 AAH01...	792	0.0
CRA 18000004926130 /altid=gi 4507409 /def=ref NP_000346.1 tran...	788	0.0
CRA 18000004926132 /altid=gi 339203 /def=gb AAA61056.1 (L02647...	786	0.0
CRA 18000005170902 /altid=gi 7657639 /def=ref NP_056564.1 tran...	561	e-159
CRA 164000136745249 /altid=gi 11968124 /def=ref NP_071979.1 tr...	554	e-156

FIGURE 1, page 3 of 4

CRA|18000005218941 /altid=gi|4572454 /def=gb|AAD23829.1|AF12128... 545 e-154
CRA|18000004926134 /altid=gi|4507407 /def=ref|NP_001053.1| tran... 128 1e-28

EST:

gi 10725490 /dataset=dbest /taxon=96...	858	0.0
gi 5936410 /dataset=dbest /taxon=9606 ...	835	0.0
gi 6888875 /dataset=dbest /taxon=9606...	726	0.0
gi 6888872 /dataset=dbest /taxon=9606...	726	0.0
gi 12258937 /dataset=dbest /taxon=960...	686	0.0
gi 10947399 /dataset=dbest /taxon=96...	680	0.0
gi 13287907 /dataset=dbest /taxon=96...	680	0.0
gi 9121897 /dataset=dbest /taxon=9606...	680	0.0
gi 13280819 /dataset=dbest /taxon=96...	680	0.0
gi 8150776 /dataset=dbest /taxon=960...	656	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

gi|10725490| adult adrenal gland
gi|5936410| adult uterus
gi|6888875| adult head_neck
gi|6888872| adult head_neck
gi|12258937| adult lung_tumor
gi|10947399| mammary gland
gi|13287907| retinoblastoma
gi|9121897| retinoblastoma
gi|13280819| adenocarcinoma cell line
gi|8150776|

Tissue Expression:

Human hippocampus

Isoform 1:

```

1 MRHLGAFLEFL LGVLGALTEM CEIPEMDSHL VEKLGQHLLP WMDRLSLEHL
51 NPSIYVGLRL SSLQAGTKED LYLHSLKLG Y QQCLLGSAFS EDDGDCQGKP
101 SMGQLALYLL ALRANCEFVR GHKGDRLV SQ LKWFLEDEKR AIDTAAMAGL
151 AFTCLKRSNF NPGRQRITM AIRTVREEIL KAQTPEGHFG NVYSTPLALQ
201 FLMTSPMRGA ELGTACLKAR VALLASLQDG AFQNALMISQ LLPVLNHKTY
251 IDLIFFPDCLA PRVMLEPAAE TIPQTQEIIIS VTLQVLSLLP PYRQISIVLA
301 GSTVEDVLKK AHELGGFTYE TQASLSGPYL TSVMGKAAGE REFWQLLRDP
351 NTPLLQGIAD YRPKDGETIE LRLVSW
(SAQ ID NO:3)

```

FEATURES:

Functional domains and key regions:

PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 2

1	75-77	SLK
2	174-176	TVR

PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 6

1	67-70	TKED
2	90-93	SEDD
3	174-177	TVRE
4	226-229	SLQD
5	249-252	TYID
6	302-305	STVE

PDOC00008 PS00008 MYRISTYL

N-myristoylation site

Number of matches: 7

1	12-17	GVLGAL
2	57-62	GLRLSS
3	86-91	GSAFSE
4	149-154	GLAFTC
5	190-195	GNVYST
6	209-214	GAELGT
7	230-235	GAFQNA

PDOC00009 PS00009 AMIDATION

Amidation site

162-165	PGRR
---------	------

SignalP results:

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	0.37	YES
max. Y	19	0.702	0.34	YES
max. S	5	0.974	0.88	YES
mean S	1-18	0.949	0.48	YES

Most likely cleavage site between pos. 18 and 19: ALT-EM

BLAST Alignment to Top Hit:

```

>CRA|108000024636236 /altid=gi|298316 /def=gb|AAB25526.1|
transcobalamin II, TC II [human, endothelial cells,
Peptide, 427 aa] /org=human /taxon=9606 /dataset=nraa
/length=427
Length = 427

```

Score = 732 bits (1870), Expect = 0.0

Identities = 376/427 (88%), Positives = 376/427 (88%), Gaps = 51/427 (11%)

Frame = +1

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted May 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Query: 31 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL 210
MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL
Sbjct: 1 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL 60

Query: 211 SSLQAGTKEDLYLHSLKLGYYQQCLLGSASFSEDDGDCQKGKPSMGQLALYLLALRANCEFVR 390
SSLQAGTKEDLYLHSLKLGYYQQCLLGSASFSEDDGDCQKGKPSMGQLALYLLALRANCEFVR
Sbjct: 61 SSLQAGTKEDLYLHSLKLGYYQQCLLGSASFSEDDGDCQKGKPSMGQLALYLLALRANCEFVR 120

Query: 391 GHKGDRLVSQLKWFLEDEKRAI----- 456
GHKGDRLVSQLKWFLEDEKRAI
Sbjct: 121 GHKGDRLVSQLKWFLEDEKRAIGHDHKGHPHTSYYYQYGLGILALCLHQKRVHDSVVDKLL 180

Query: 457 -----DTAAMAGLAFTCLKRSNFPNRRQRITMAIRTVREEILKAQTPEGHF 597
DTAAMAGLAFTCLKRSNFPNRRQRITMAIRTVREEILKAQTPEGHF
Sbjct: 181 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFPNRRQRITMAIRTVREEILKAQTPEGHF 240

Query: 598 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 777
GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT
Sbjct: 241 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 300

Query: 778 YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLGASTVEDVLK 957
YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLGASTVEDVLK
Sbjct: 301 YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLGASTVEDVLK 360

Query: 958 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 1137
KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI
Sbjct: 361 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 420

Query: 1138 ELRLVSW 1158
ELRLVSW
Sbjct: 421 ELRLVSW 427
(SEQ ID NO:6)

HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	829.9	8.6e-246	2
CE00052	CE00052 lymphocyte_transmembrane_protein_KAP	3.2	2.9	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11 [.]	1	11 [.]	3.2	2.9
PF01122	1/2	1	142 [.]	1	143 [.]	296.0	4.6e-85
PF01122	2/2	143	376 .]	197	450 .]	531.8	4.8e-156

FIGURE 2, page 2 of 4

Isoform 2:

```

1 MRHLGAFLEFL LGVLGALTEM CEIPEMDSHL VEKLGQHLLP WMDRLSLEHL
51 NPSIYVGLRL SSLQAGTKED LYLHSLMLGY QQCLLGSAFS EDDGDCQGKP
101 SMGQLALYLL ALRANWHDHK GHPHTSYQQY GLGILALCLH QKRVHDSVVD
151 KLLYAVEPFH QGHHSVDTAA MAGLAFTCLK RSNFNPGRRO RITMAIRTVR
201 EEILKAQTPE GHFGNVYSTP LALQFLMTSP MRGAELGTAC LKARVALLAS
251 LQDGAFQNAL MISQLLPVLN HKTYIDLIFP DCLAPRVMLE PAAETIPQTQ
301 EIISVTLQVL SLPPYRQSI SVLAGSTVED VLKKAHELGG FTYETQASLS
351 GPYLTSVMGK AAGEREFWQL LRDPNTPLLQ GIADYRPKDG ETIELRLVSW
(SAQ ID NO:4)

```

FEATURES:

Functional domains and key regions:

PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site
198-200 TVR

PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site
Number of matches: 7

1	67-70	TKED
2	90-93	SEDD
3	147-150	SVVD
4	198-201	TVRE
5	250-253	SLQD
6	273-276	TYID
7	326-329	STVE

PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 7

1	12-17	GVLGAL
2	57-62	GLRLSS
3	86-91	GSAFSE
4	173-178	GLAFTC
5	214-219	GNVYST
6	233-238	GAELGT
7	254-259	GAFQNA

PDOC00009 PS00009 AMIDATION
Amidation site

186-189 PGRR

PDOC00428 PS00468 COBALAMIN_BINDING
Eukaryotic cobalamin-binding proteins signature
165-178 SVDTAAMAGLAFTC

SignalP results:

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	0.37	YES
max. Y	19	0.702	0.34	YES
max. S	5	0.974	0.88	YES
mean S	1-18	0.949	0.48	YES

Most likely cleavage site between pos. 18 and 19: ALT-EM

BLAST Alignment to Top Hit:

```

>CRA|108000024636236 /altid=gi|298316 /def=gb|AAB25526.1|
transcobalamin II, TC II [human, endothelial cells,
Peptide, 427 aa] /org=human /taxon=9606 /dataset=nraa
/length=427
Length = 427

```

Score = 793 bits (2026), Expect = 0.0
Identities = 399/427 (93%), Positives = 399/427 (93%), Gaps = 27/427 (6%)

Query: 1 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPMDRLSLEHLNPSIYVGLRL 60
MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPMDRLSLEHLNPSIYVGLRL
Sbjct: 1 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPMDRLSLEHLNPSIYVGLRL 60

Query: 61 SSLQAGTKEDLYLHSLMLGYQQCLLGSAFSEDDGDCQKGKPSMGQLALYLLALRAN----- 115
SSLQAGTKEDLYLHSL LGYQQCLLGSAFSEDDGDCQKGKPSMGQLALYLLALRAN
Sbjct: 61 SSLQAGTKEDLYLHSLKLGYYQQCLLGSAFSEDDGDCQKGKPSMGQLALYLLALRANCEFVR 120

Query: 116 -----W-----HDHKGHPHTSYQYGLGILALCLHQKRVHDSVVDKLL 153
W HDHKGHPHTSYQYGLGILALCLHQKRVHDSVVDKLL
Sbjct: 121 GHKGDRLVSQLKWFLEDEKRAIGHDHKGHPHTSYQYGLGILALCLHQKRVHDSVVDKLL 180

Query: 154 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFPNRRQRITMAIRTVREEILKAQTPEGHF 213
YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFPNRRQRITMAIRTVREEILKAQTPEGHF
Sbjct: 181 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFPNRRQRITMAIRTVREEILKAQTPEGHF 240

Query: 214 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 273
GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT
Sbjct: 241 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 300

Query: 274 YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLAGSTVEDVLK 333
YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLAGSTVEDVLK
Sbjct: 301 YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLAGSTVEDVLK 360

Query: 334 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 393
KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI
Sbjct: 361 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 420

Query: 394 ELRLVSW 400
ELRLVSW
Sbjct: 421 ELRLVSW 427
(SEQ ID NO:7)

HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	906.3	8.6e-269	2
CE00052	CE00052 lymphocyte_transmembrane_protein_KAP	3.2	2.9	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11 [.]	1	11 [.]	3.2	2.9
PF01122	1/2	1	115 [.]	1	115 [.]	241.3	1.4e-68
PF01122	2/2	117	400 .]	145	450 .]	660.5	8.7e-195

1 ATATGTATGG GAAATATGCT GTCTTCCTAT TCCTACTCCC CCACCCCTCTA
51 GCACTGAGTC CAGGTAGGTA GGCAGGGGGG TGTCTCCCTC CTTTACTTCG
101 ACACCCCTAAC TACCTTGGGG ATCAGAAGTG ACTCTCTGGA AGGATGCTGC
151 TGCTTCTCAC CAGAGGCTGA CGATAACGAA GGCTATCCTC CATGGCCACC
201 TCCTCCAGGC TGCCTTCCTG GAAATAGGAA TCATAATAGT TGTACTGGA
251 AACAGGCAGA GGGTTGGGGG AGCCAAGGCA GTCCCACCCA GGACCAAGGT
301 GGCTCCATTG CACACACTTC ACCATGACTC CCCTGAAGGT CCAAACGTGC
351 GGTTCGCGG AAGTTGGGCT CCCCACTGGC CTCCCTCCTT CCTCAGAACC
401 TCCAGGGGTG CTCCTCCTAG TGGCCACATC CAGCCTTTCT GACTGGACAA
451 CCTATCATTT AAAATTTTCA AGTAGTTCG TAAACAGACA CACGTTGCTG
501 TATTTATTTA TGTCAAGGGC TTGGTTTGTG ATAAGTCAGG CTCAAAAAGA
551 TTGTCTTAAA AGAGTGAACC TTGGCAATTT ACCATAAAAT AATTGCAATG
601 CAGATTGTGC ATGGAATGA TTGGAGATAT TTAAGGTCA TAGTGTCTTC
651 ACAAATTGAG CTGAAGGGA ACTGTTAGGA TGATCTTGCC TAACCTCTC
701 ATCTCACACA GGAAGAACTA TTTTAAACTC GAGAGGTAA GTACCTGGC
751 CAAAGTCACA CAGCCACCAC TAGTTAACTC GTATACATTG ATTCTCCTGT
801 GGGGCTGGGC AGATGAGGAA TCTTTTGTTC TCTCCCTGT TGCAGAGAT
851 TTTTCTTGAG GTTACTTTCC GAGTCTGGC AAGTACCCCT GCTTCTGGTA
901 GCTTGTGTG TGTATTCAAT CTCATTCTTT TTATTTTATT TTATTTTGA
951 GACAGGGTCT CACTTTGTCA CCCAAGCTGG AGTGCAGTG TGTAATCTTG
1001 GCTCACTGTA GCCTCCACCT CTTGGGTTCA AGCGATCCTC CTGCCTCAGC
1051 CCCCCAAGTA GCTGGGATTA CAGACGCTG CCACCACGCC AGGCTAATTT
1101 ATGGTTTTTT GTATGTGTTT TTTGTGTTTT TGTAGAGACA GTGTTTCCCC
1151 ATGTTGCCCA GGCTGGTCTC CAACTCCTGA GCTCAAGTA TCTGCCCGCC
1201 TCAGCCTTTC AAAGTGCTAG GATTACAGGT GTGAGCCACC GTGCCCGGAC
1251 TTAATCCCAT TCTTTAACTT GTTTTGTGTT GTCTCTCCA GGAGGCTCCC
1301 AGCCCTTTTC GATTGGTTGA GAAAAGTGGC CTGGCTGGTC TGGGGCCAGC
1351 AGCACCCACC CTCCCCTCAA TTGCCCAACT CCCCCCCCCA CCGAAGTGCC
1401 CAACTCCCCC TCCCCAACTG CCCAACTCCC CCACCCCCAC AATCCCTCC
1451 CGCCACAAC T GAGGGAGGCG GTGCTGAAAA ACAGCTGACT CCAGCAATGC
1501 TGCTCACGTG ACCACTGCAG CTGCAGCTCC CGTTCCACTC CTGTCTCTGG
1551 GCTAGGTGGG CACTACCAGG GGCTCCTTTG GTAAGGAGTA CCGGGTAGGC
1601 ACCCGGTCCT GCCAATCCAC CACTGGAACA GCTGGGGGGA CAGCAGACAG
1651 GCACGGTCGG ACAGACTTGA CAGATCAGGC ATCAGGCCCT CTGCGCTGGT
1701 CCGGGGCTCT TTAAGCAGGA ACGTGAATGG CCTCAAGATG TCTCACATGG
1751 TCCCACTAGC CCTCCTCCTC CTTTGTGTTT CTACCTCCAG GAGGGCTGCT
1801 CTGCCCTTCC TTCCTCTGTT CTTTGGCCTT ATGTTCCCG CCACCACAGG
1851 CCTTCCCCCG CCCCACCCCT CTGCAGACTT AGCCGTGCAT TGCAGGCATG
1901 GAGGATTAAT CAGTGACAGG AAGCTGCGTC TCTCGGAGCG GTGACCAGCT
1951 GTGGTCAGGA GAGCCTCAGC AGGGCCAGCC CCAGGAGTCT TTCCCCGATC
2001 TTGCTCACTG CTCACCCACC TGCTGCTGCC ATGAGGCACC TTGGGGCCTT
2051 CCTCTTCCTT CTGGGGGTCC TGGGGGCCCT CACTGAGATG TGTGGTGAGT
2101 AACTGCGCTC TATCCTGTGC CTCTTTCCTC CTGGGTCCCT AGTGGGGTGG
2151 CTAGGCATA GATGAGGGA ACTTACCTGC CCTTCTAAGC TCCCATAGCA
2201 GTTTGGGCTT AGCTGGACCT CAGCATTTAA CACATCCTAT TGTGATTGAT
2251 TATATGTTTG ACTCCTCACC AGACAAGATC TCCGTTAATT CAGTCATTCTG
2301 TTCACACATT CATTCAGCGC ATACTGAGCC TTTTCTGTGT CAGGCCAGT
2351 GTTAGCCTTT GGGGAACGTG CAAAGCATGA GACAAGTCTA ATCCCTGCCA
2401 TCCTAGAGCT TATGTTCTAG GGAAGGGGGA CAGACAAAAG AAATGGTTAG
2451 GTGCTCCAC CTGAAATCTC AGCATTTTGG AAGGCTGAGG CGGGAGGGGA
2501 GGATCGCTTG AGCTCAACAG TTCAAGTCA GCCTGGGCAA CATAGGGAGA
2551 CCCCATCTCT ACAAAAATA AAAAAATTA AAAAAATAGT GGGCATGGGG
2601 AAGACTTTCT GAAGACCAAG AGGACACATG GGAGCTGAAA CTCGAAGGAA
2651 GAAAAGGAGC TGGCAGGAAA GGAGTGGGGG ACACACATTC TAGGCAGCAG
2701 GAAGTGAGCC TTCGGAGGTC CTGCCTGCTC CAGCTCTGTG CCCCAGGGG
2751 TCTCTTGGAG CACAGTCTCC TGGGACCTGT CTATGAGTCT GAGCTTAGAG
2801 GCTCAGGGCT GCTCCTTCAG ACAGGAGGCA GAAGGCAGAC TTTGGGAAC
2851 TGGGGCCGCC CACGCGCCTT TTCTCCTCCT CTGCACCTAG GATTACGTTG
2901 AGCAATACAC TTTACCCCC ATGGTCTCTT GAGACCCTGG GGAACCCCTG
2951 AGAGGTGGGT CAGTCATGT CCAGGTGTC AGTGAAGAAG TCGAGGGTTG
3001 GAGGGGCTGA GTGACCCACT CAGGGTGCTC CACCTTTTCC AGAGCTTTGC
3051 TGAACCTAGT TTTTAGAAGT TGAAGCCTCG TTTGTTTTTC TTTTGTGTTT
3101 TGTTGAGAGA GGTTCCTCCT CTGTTGCCCA GGCTGGAGTG CAGTGGCAGG
3151 ATCTTGCTCT ACTGAGCCT CTGCCTTGTG GGTTCAGTG ATTCCCCAC
3201 CTCAGCCTCC CAAGTAGCTG GAGACTGCAT GTGCATACTA CCATGCTTGG
3251 CTAATTTTGG TATTTTGGG TAGAGACAGG GTTTCGCCAT GTTGGCCAGG
3301 CTGGTCTCGA ACTCCTGGGC TCAAGTGAAA CTCTTGCCCT GGCCTCCCAA
3351 ATTGCTGAGA TTACAGGCGT GAGCCACCGT GCCCGGCCAG AACTCCAAGC
3401 CTCTCATCTG TGTTCATAA ATGCAATCAG ACACCTCAGG TCTGGGCCCA
3451 GGAACCCAG CTCTTGGTTC ATGTCCGGAC AGTCCCGAGG GGAGTTCTGG
3501 GTTCAACCAG CAAGAGCTCT TCCTCCTGGC TGATCTGGTC CTCAGCCTTG

FIGURE 3, page 1 of 22

3551 GACAGTTAGT CCATTAACCT GACCCACAG GAGCCCAAT CCCTGGGGT
3601 CTGGGGAATC TTGAACTGGG GTTGGGGTG CAAATATCTG CACTGAGTCA
3651 CTTAATTGCA CCCAGCCTCA TTCCTTTATC TGTAAGTGG GCTAAGAATG
3701 CTCCCCTGCC TTCCTCTCG GTGTAGTACG AGGAAGGATC CCATGACACC
3751 TGCTCTCCCA GTTTAAAGCT CTATATGTAT GTTGTGAAAT TGACAGGGAT
3801 CGCTGCACAA ACGCTAATGC AAAGTGGGCT CCTGTGCTTC CTTTCTCTT
3851 TCTTCTCTT TTTTTTTTT TTAATTTTCT TCTAGAGATG AGGTCTCACT
3901 ATATTGCCCC GGGTTGGTTT CAAACTCCTA GGGTCAAGCG ATCCTCCAC
3951 CTGGCCTCC CAAACTGCTG GTATTACAGG CGTGAGCCAC TCTGTCTGGC
4001 TCCTATGCTT GTGAATGTCA ACAGCAATCA GCCCTTAGCT GGCAGGGCTG
4051 GGTGGTAGG GCGAGAGCTC ACCCAAGGCT GCTTTTATTA CCCTGCGTGA
4101 ATCTGCCTGG CCCCTTCTT CTAAGGAGGT TGCTCTGTGG TTGTCACTCT
4151 CTCCCTTTAC AGCTGGATCC TGATCTTTCA GTTCTAACC CTGTCTGAC
4201 TCATCGTGCT GGAAGTGAGA GCCCGGGGTG AGGTCAAGGA ACTCCCTTGC
4251 CGGTTTCAAG AAAAGGGAAA AGGAAAGAGA GGTGAGGAGG GGGGCGATG
4301 ACCAGAGAGA CACAGGCTGA GAGAGACTGA GACAGACCCA GAGAGCCTCA
4351 CACATTGAGT GACAGAGACG GAGAAATGGA GATAGGCACC AAAAATGGT
4401 TCTCAGTGAC AGAAAGGGAA AAAAGCAACC CCCAGTCTC TCTTAACATC
4451 TGGTGAGAAA CAGCCATGT GCTTGGTCT GGGCCACAC AGCAAGGAT
4501 TATGTAGGGT TTCATGCTGG TGGATGGTCA CCTTATAGCA ACAGGTATCT
4551 GGGGCTGTGG GGAACACAGA CACGAGGTTG TGGGACCCAG ACCACAGAG
4601 ATGGAGCTGT TCTAGGAGCT CTGGTCTCG TTCTGGTCCC CTGGGATATG
4651 GCACAGTGAA GGCACCATC AGGCAGCTGG AGCCAGCAG CAACTGGGAG
4701 GCAGTAAACA GGGACCGAAA GTGCAAGGTT ACCTCCGAGG CAAACTACTC
4751 TAAGCTACCC TGTGCTGAGC TCAAGTCCCT TGGAACATC CTAAGGCTT
4801 CCGCTTCCAG AGTGTGAGG TATTTTCGTT GCACAGCTTC GAATAAATCC
4851 CACAGCAACA GGTAAACGGC TGCAAGCTGT GACTGTTTC TAAGAGCTCA
4901 TCTCACAATC TCAGGTCCTC TTCATTAAAG CAGAGATGGC AGGAAAGGCG
4951 TTATTTTGAG ATCTGCATGG AGGAAGTTCA CCAGGCAGCC TCAATTCAAC
5001 AGCTGGAAGT TTGCGTTGTT TGGAAATTTG ATGTGTAACA CGTTCTGCAT
5051 GTGGGCTGAT GTTTTTGTAA ACGGGTAGCA CACACATTCA GCAGGGCACC
5101 AAAGAGCGGG GGCTTTGCAG TTAGGTCCAT CCTTGGCTCT GCAGCCTTGT
5151 GTAAGACATG ACACGACTTT GAACTTCTGT TTCCTCTTCT GTGCAAGCA
5201 ATGATGACAG TATCTACATC ACAGGACTGG CATGAGGACC AAGTGAGATT
5251 GGGCAAGGTG CCCGGGCACA CCACTCTCAC TGTCACTGCT GATGGGCGA
5301 GTGGTTGCC TGGCAGTAGCA TCCTCTATCT TCAGCCACC ACCTCTCTTG
5351 CTGGCTCACT CCAACTGCTC TTTAGAGATA CACGCTTCCC CTCTTTCTC
5401 CTCCCACTGC CTTTCAGTAT GGCTGCATTT CCCCTGCAA GTTGGTGTGT
5451 GCTGGGTGGA GGTGGGGGTG AGGACATGTA TTCTCTGGAG AAGGCCCTGG
5501 TAACGTCAA GCACCTTCTT GCTGGTGGCC TGGCCCTGTG ACCTCATTTG
5551 TACCATTTTC TTTTCTAAGA AATACCAGAG ATGGACAGCC ATCTGGTAGA
5601 GAAGTTGGGC CAGCACCTCT TACCTGGAT GGACCGGCTT TCCTGGAGC
5651 ACTTGAACCC CAGCATCTAT GTGGGCCTAC GCCTCTCCAG TCTGCAGGCT
5701 GGGACCAAGG AAGACCTCTA CCTGCACAGC CTCAGCTTG GTTACCAGCA
5751 GTGCCCTCTA GGGTATTGCC ACACTCTCTT TTTCCATGTC TTGCTCCACA
5801 TACTAAGAGA TGGGAAACTT GGGTACTAGT TTGGGCCTGT CACCACTTTG
5851 TGGGCAGACC TTAGGCAAAAT TTTCTCCATC TATAGAATGG AGGACCTTTG
5901 TCCATCTATA GAATGAAGGG GTTGGTTGGA TTAGATCAGA GATGCTAATG
5951 CAAGGCTCCT TTTGCTACTA CTGTCCATCA TGTGTCTGAG GCAGACATAA
6001 CTAATCCGTG ACTATACTCT TTGATGATGA GCCCAGGAGC AGCATCTGAC
6051 TCTATGCTCC CTTAGTGTGC CTGAGGCAGA TATCACTAAT CGATGACTGC
6101 AGTCTTCTAC ATTGAGCTTA GAAGCAGCAT CTGACTCTGT ATGCTCCCT
6151 CCCATGCATG AGGCAGACAT CAGTAATCCA TGACCGCATT CTTTCATACT
6201 GAGCCAGAA GCAGCATCTT TTCTTTCTT TCCTCTCACT CTGTTGCCCA
6251 GGCTAGAGTG CAGTGGCACA ATCTGGCTT GCCCAACCT CCAATTCCCG
6301 GGTTCAAGTG ATTCTCGTGC CTCAGCCACC TGAATAGCTG GGATTACAGG
6351 CGTGTGCCAC CATGCCAGC TGATTTTGT ATTTTGGTA GAGATAGGGT
6401 TTCACCATGT TGGCCAGGCT GGTCTTGAAC TCCTGACCTC AGGTGATCCG
6451 CCTGTCTTGG CTTCCCAAAG TGTGGGATT ATAGGCATGA GCCACTGCAC
6501 CAATCCAAA GCTCCCTCT TGAGCACTT TGTGCTCCCT TTTCAAGAGG CATCACAGAG
6551 AGGCCTGTTT TGGGGTTTGA ATGAGAGGCG AAGAATCAGC CATGGAGTGC
6601 CTCTTTCTCA GACTCCCTCT TGAGAAGTGG GTGCAGGGGT GGAGAGAAAA
6651 GAAGACTAGG CATAGTGGCT CATACTGTA ATCCCAACAT TTTGGGAGGC
6701 TGAGGCAGGA AGATTGCTTG AGCTCAGGAG TTTGAGACCA GCCTAGGCAA
6751 CATAGTGAGA CCACATCTCT TAAAAAAG AAAAAGAAAA AAAATGAGCC
6801 AGGTGTAGTG ACTCATGCCT GTGGTCCCA CTCTCCGGA GGCAAGGGT
6851 GGAGGATCTT TTGAGGCTGA GAAATCGAGG CTACAGTGAG CCATGGTGGC
6901 ACCACTGCAC TCCAGCCTGG GAGACAGAGA GACCCTATCT CAGTAAAAAA
6951 AAAAAATAAA AATATGGCTG GGTGTGGTGG CTCACGCCTG TAATCCAGC
7001 ACTTTGGGAG GCCAAGGTAG GTAGATCACA TGAGGTTAGG AGTTCGAAAC
7051 CAGTCTGGCC AACATAGTGA AACCTGTCT CTACTGAAAA TACAAAAAT

FIGURE 3, page 2 of 22

7101 TAGCCAAGGG TGGTGGTGGG CAACTGTAAT CCCAGCTACT TGGGAGGCCG
7151 AGGCAGAAGA ATCGCTTGAA CTCGGGAGGC GGAGGTTGCA GTGAGCTGAG
7201 AACATGCCAC TGCACTCCAG CCTGGGCAAC AAGAGCGAAA CTCTGTCTCA
7251 AAGAAAATAA ATAAATAAAA TAAAAAATA AAAAAGGAGG GGGCATATGG
7301 GTGAAGTATG GACAAAAATAG TGGGGCAGGC ACAGATGATC TGGACACAGG
7351 AGCCCTTGGA GTTTATTCTT GAATCTAACT GTTCATCTTT ATTAAATATT
7401 TGTGGCATAAC ACCTCACAAC AACATAGCCA ACACACCTCC TTTTGGAGCT
7451 TTTATCGAAG TTTCCCACTG TTAAGATTTT TTCCCGCTTT GTGATGCGGG
7501 TGGGGTGGGT GCTGTAAGCA GGCTTACGGG GTGGCAGTTT CTCACAAAGG
7551 CATTAACTGG CCTTGTCCCTA GGTCTGCCTT CAGCGAGGAT GACGGTGACT
7601 GCCAGGGCAA GCCTTCCATG GGCCAGCTGG CCCTCTACCT GCTCGCTCTC
7651 AGAGCCAACGT GTGAGTTTGT CAGGGGCCAC AAGGGGGACA GGCTGGTCTC
7701 ACAGCTCAAA TGGTTCCTGG AGGATGAGAA GAGAGCCATT GGTGAGCAGA
7751 CACCATCCGC TGGGGGTGGG GAGCAGCTGG GAGGGCTCAT CAGATGATAT
7801 TCTCCAATGA GAATCAGAAC TTTGGGTTTT CTCCCAGGC GTCTTTCCCA
7851 CCATCCATTC TGCCCATCTC ACTGCCTACG TAGAGGCTCG AACCTGTCCC
7901 CATAGCCATC CTTGACCCAG CTTTTCCTCG GCTGCACACA TACTATTGAC
7951 AGGTGTGTTT CGTGGTTTTT TGTTTTTTGT TTGTTTGT TTGTTGAGTT
8001 GGAGGTTTGC TCTTGTGCTC CAGGCTGGAG TACAATGGCG CAATCTCAGC
8051 TCACCCGAAT CTCTGCCTCC TGGGTTCAG CAATTCTCTT GCCTCAGCCT
8101 CCTGAGTAGC TGGGATTACA GGCATGCGCC ACCACACCCA GCTAATTTTG
8151 TATTTTTAGT AGACGTGGGG TTTCTCCATG TTGGTCAGGC TGGTCTCGAA
8201 CTCCTGACCT CAGGTGATCC GCTTGCCTTA GCCTCCGAAA GTGCTGGGAT
8251 TACAGGCATG AGCCACTGCG TTAGGCCCAC TGACAAGCCT TGTATTGGCT
8301 AGCCACCAAG ATTGACTTGA TTATCCACCT TCGGGACAAC TGGACAGCCT
8351 GCTTATGACT TACGCCATAG TCTGTCTCTA CTAGCTCTCC TGCCCTGACT
8401 TGACCCAGCA TACAACAGCC AGAGCCAGCC TTTTCAATAT AAACCTGATC
8451 TTGCTGGCAC TGCTTAAACC CTGCAGGGGC CTCGCACTGC TCCATGGCCC
8501 AGCCTGTCTA CCCTTACCTT CTGCCCAGGC TCTGCTCATC CATTCTCTGC
8551 CTCCACACA CCTGCCCTCT GTGGGCTCCA GCCATACCAT CTCTCAACTC
8601 ATAAGCCAGT TTTTTCATAC AGGCTCCCTC CATCTGGACT GGCTTCCCTG
8651 CGTGCACTT ACCTCTGCTC TACCTTTGGC TCTGCCTCCA CCCATCCTCA
8701 GCCGTCTCCA GCATTACCTC CTTGGAGAAT CCTGCCTTGA CTTCCAGGCC
8751 ACCCAAATAT CACTACTTGG TCTGCATTCT CGTTGCAATT GCAGTCGCAT
8801 GAGCAATTGC TGTGGTTGAG GCCCGAAGTG CGCAAGTGCC TGTCTGCCAT
8851 GGTCTCTCTG CTCTCTCTAA GCACAGTGCC TGACACACAG TGAGACCTCA
8901 GCACGTATGG GCTGAGGCAA TGAAGGAATG AAGGATCCCA TGACCCAAAA
8951 GAGCCTGTGG GAAAGTGCA GCGAGGGTCC CAGGTGCTGG CGGGGCTGGC
9001 TGCTGGGTGG GGGCAGAGAG GCAACCCCTC TGTTTTTTTC CCTCTCAGGG
9051 CATGATCACA AGGGCCACCC CCACACTAGC TACTACCAGT ATGGCTGGG
9101 CATTCTGGCC CTGTGTCTCC ACCAGAAGCG GGTCCATGAC AGCGTGGTGG
9151 ACAAACTTCT GTATGCTGTG GAACCTTTCC ACCAGGGCCA CCATTCTGTG
9201 GGTGAGTAGG TCAGACCGTG CCAAGGCCAG GCTGGCACTC CCTCAGTCCC
9251 CAGGTCTCTG CTGATGACGT CCATACCCCTG GCCCCACAC TCACCTTTCC
9301 TTGGGGCTCC TCCGAATCAA GTCCCTTAGG GACGAATTGG CGAGGGCTCA
9351 TGGGTGATGC TCCAGCTGTG AGCCAGCTTT GGAGCTGGTA GGTGGATCTC
9401 TTGAGGCCAG GAGTTCAAGA CAACGTGGTG AAACCCCATC TCTACTAAAA
9451 ATAAAAAAGT TAGCCGGGCA TGGTGGCACA TGCCTGTAGT ACCAGCTACT
9501 CGGGAGGCTG AGGCAGGAGA ATCACTTGAA CCTGGGAGGC GGAGGCTGCA
9551 GTGAGTGGAG ATCGCACCCAC TGCCCTCCAG CCTGGGCAAC AGAGTGAGTG
9601 AGACTCTGTC TCAAAAAATA AAAAAATAA TAAACTCCC CTAGTGATTCT
9651 CAATGTGCAG CTAAGTTTGG AAATAGGTGG TATGGGGTCA AGTCCCTTTG
9701 GGCTCTCCCTC CTCCAGTCCCT TCTCCCTAAC CTCTAGCCCT CAAGTTGCAG
9751 AGTGATCAGC CAAACCAGTT TGCCAGAAA TGAGCAGTTT CCTGGGACAC
9801 AGGATTTTCA GAGTCCAGAC AAGGAAAGTC TTGGGCAGAC CAGGTTGAGT
9851 TGGTGCCTTT AGCTGATCTG ACCATGTTGC CCTTCTTCTC CAAGCCCTCC
9901 TGTGGTTGTC CATAGCTACA AGGGCCTGAC CCTCAAGCCC CTGCCTGTCC
9951 TGGCCCTTTT GGCTCTCCAG CTCATTGCAT GTTCTGTCCC CCACTTCAAG
10001 ACACAGCAGC CATGGCAGGC TTGGCATTCA CCTGTCTGAA GCGCTCAAAC
10051 TTCAACCTCTG CTGGGAGACA ACGGATCACC ATGGCCATCA GAACAGTGCG
10101 AGAGGAGATC TTGAAGGCCC AGACCCCGA GGGCCACTTT GGGAAATGTCT
10151 ACAGACCCCC ATTGGCATTA CAGGTGGGAA AGAGACCCCTG GAGCCATGGC
10201 CACCTTGGGG AACAGTCGGG TGGAGTGGTC AGGTGCTGGA ACACCTAGCC
10251 CCTCCCTGCC GGCTGACCTC CTCTCTCTCT TCCTCACTCT ATCACCAGTT
10301 CCTCATGACT TCCCCATGCT CTGGGGCAGA ACTGGGAACA GCATGTCTCA
10351 AGGCGAGGGT TGCTTTGCTG GCCAGTCTGC AGGATGGAGC CTTCCAGAA
10401 GCTCTCATGA TTTCCAGCT GCTGCCCTGT CTGAACCACA AGACCTACAT
10451 TGATCTGATG TTTCCAGACT GTCTGGCACC ACGAGGTAGC CCAACTTTTT
10501 GTGGAAGCAC AGCCCTTTAC AATCTGCTGC GCACCCATTG ACGTCCCAGT
10551 GAGGGGAGGT TGCTTCATCC TGATTTGCTG AGTCAGCACA AGTTTGTGGG
10601 TGTGCATGGG ACACAGTAGC CAAAATGTGG TCATAGCTTC TAGAAGCTCA

FIGURE 3, page 3 of 22

10651 CAGTGTGGGG AGGAAGACAG TAAATGGAGA TCCCTGGGCA TATCGCTTGT
10701 GTGATACCCA GTACAGAAAT GTTTGGATGG ATGGATGGAT GGATGGATGG
10751 ATGGATGGAT GATGGATGG ATGAGGAGAG ACACATTTTG GTTAACTCTA
10801 ATACAACATG ATAAGCCCCA GTAGCAGCAT GATCCAGGCT TTCTCTGAGA
10851 GAGGGTCTGA GGACGTGACT GGGATTTGCC AATTAAGAAT GGAGAAAGAG
10901 GCCAGGTGCA GTGACTCATG CCTGTAATCC CAACACTTTG GGAGGCCGAG
10951 GCGGGTGGCT CACCTGAGGT CAGGAGTTCG AGACCAGCCT GGCTAACATG
11001 GCGAAACTCC ATCTATTAAA AATACAAAAA AGTAGCTGGG TGTGGTGGCG
11051 AGTGCCTGTA ACCCCAGCTA AGCTACTCAG GAGGCTGAGG CAAGAGAATC
11101 ACTTGAACTT CAGAGGTGGA GGTTCAGTGT AGCCAAGATC ATGCCACTGC
11151 ACTCCAGTCT GGGTGACAGA GTAAGACTAT GTCTCAAAAA AAAAAAAAAA
11201 AAATGGAGAA GAAGGAAGCT GGACATGGTG GCTCGTGCTT ATAATCCTAG
11251 CACTCTGGGA AGCTGAGGCA GATGGATTGC CTGAGCCAG GAGTTTGAGA
11301 CCAGCCTGGG CAACATGGTG AAACCCTGTC TTTACTAAAA TACGAAAGAT
11351 TAGCCAGGCA TGGTGGTAGA CACCTATAAT CCCAGCTACT AGGGAGGCTG
11401 AGCCACAAGA ATCACTTGAA CCTGGGAGAC AGAGGTTGCA GTGAGCCGAG
11451 ATCGCGCCAT TGCATCCAG CCTGGGCGAC AGTGTGAGAC TCTGTCTCCA
11501 GAAAAACAA GAATGGATAG AGTGGAGCCA AGAAGAGGCA GGAAGAACAA
11551 AGACACAGAG GTGCACAGAG TTTGGGGGAA TTTGAGGAA TGGTCTTGCA
11601 AAAGAGTGGG ATCTGGGAGA ATGAGTGGGA GTGGAAAGCA GATGAATGAA
11651 GAGAAGGTGA GCGCATCAGG GTAACAGAGA TCGTGTGTA ACAAAATGCAT
11701 GTTCTAGGAA GAGCCCTCTG GAGTGCTAGG TGCCAGAGAG GTGGGAGGAA
11751 GGATACTGGA AGCAGAGAAA CCAGTGAGG GCCTGATCTT GGGTGGTGGG
11801 GAATGAGGGA CAGGGGAGGC CGGGATGGAA GCCAGGTGGT GGGGAATGAG
11851 GGACAGGGGA GGCCGGGATG GAAGCCAGGT TTCAGCTGAG CAGGTGGCGG
11901 TGGCATTGAT GGAGATGAGG ACATGGGGAA GGACAAAGTC CAGGTGTCCT
11951 TGAGGGAAGA CAAGAAGACA AATAATCCAG GCTCTCTGTC CTCACACCAG
12001 CTGCCCGCCC CTTTCTTCTT GGCACAGTCA TGTGGAACC AGCTGCTGAG
12051 ACCATTCTCT AGACCCAAGA GATCATCAGT GTCACGCTGC AGGTGCTTAG
12101 TCTCTTGCCG CCGTACAGAC AGTCCATCTC TGTCTGGCC GGGTCCACCG
12151 TGGAAGATGT CCTGAAGAAG GCCCATGAGT TAGGAGGATT CACGTGAGAC
12201 TCCCACTCC CAGTCTCAC CCCACCAAC CTCACATGCC TGATAACAGG
12251 GTCACAGAAA AGACGGGGAA CAGAGGAGAG GGTTCCTCG GGAGAGACAC
12301 TGGCCCTGCT TCTGCTCTA CCTGCTCAGC TCCTTTCTTG CCCACGGTGT
12351 TATGGAACA GGGAGCCATA GGCCAGCATT GTCAGTGA GAGCAGGCTT
12401 TGGAGGCGA GCCCCAGT TGGAAATCCA ACTCTAACA GCTAGGTTCC
12451 AGGTAGGCAC CCACAATTCA CCGAGGAGAA CAGTTGTGCC CCTTCCCTGC
12501 AGGGCCAGTG TGAAGAGTCC AGGAGTTAGT ACACATAGAG ATAGTGGCAT
12551 GTGCTTTTAT TATGTGCAAG GTCCAGCACA TAGCAAGCGC TCAACACAGC
12601 GTTGCTTTCA TCAGAGTAAG AACTGTTTTT TGTGTTGTTG TTTGTTGTTT
12651 TTTAAGAGAC AGGGTCTCAA TCTTATCACC CAGGCTGGAG TGTAAATGTG
12701 CAATCACGTC TCACTGAGT CTCGAACCTT GGGGATGAAG CAACCCCTACT
12751 GTCTGCTCTC AGCCTCCCAA ATAGCTGAGA CTATAGGCAC GTGCCACACA
12801 ACCCTGGGTA ATTTTTTTTT TTTTTTTTTT GAGATAGGGT CTCTGCTGT
12851 TGCCAGGCT GGTCTCAAAT TCCTGGCCTC AAACCATCCT CACACCTGAG
12901 GCGCTCAAAA TATTGGGATT ATAGGTGCGA GCCATCATGC TCAGCCAGAA
12951 TAATAACTGG TTTTTTTTGT TTTTTTTTTG AGACAGAGTC TCACTCTATT
13001 ACCCAGGCTG TGAGGCCCA ACTCGTGTG GTGTATTTGT TTTATTTTAT
13051 TTATTTATTT ATTTTCGAGC AGAGCCTCTC TCTTTCACCT AGGCTGGAGT
13101 GCAGTGGCGC AATCTCGGCT CACTGCAACC TCCGTCTCCT GGGTTCAGT
13151 GATTGCTCTG CCTCAGCCTC CTGAGTAGCT GGTGCTACAG GCGCGTGCCA
13201 CCATGCCCAG CTAATTTTGT TATTTTGTAG AGAGACAGGG TTTTACTATG
13251 TTGGCCAGCT GGTTCCTAAC TCCTGAACCT GGGTGTCTG CCTGCCTCGG
13301 CCTCCCAAAG TGCTGGGATT ACAGGCATGG GCCTCCGTGC CCGGCCATGT
13351 ATTTATTTAG GCAAGGTCTC TCTCTGTTAT CCAGGCTGAA GTGCACTGGC
13401 ACATTCTATG CTCACTGCAG CCTCAAATTA TCCAAGTAAC AGGGACTACA
13451 GGCATGCACC ACCACACCCA TCTACTTTTT TTTGAGATGG AGTCTCCCTC
13501 TGTGCGCCAG ACTGGGTTGC AGTGGCACAA TTTCACTCA TGGCAGCATC
13551 TACCTCCCAG GTTCAAGCGA TTCTCCTTCC TCAGTCTCCC GAGTAGCTGG
13601 GACTATGGGC ATGCACCACC ATACCTGGCT AATGTTTATA TTTTGAGTAG
13651 AGATGGAATT TTGCCATTTT GGCCAGGCTG GTCTTGAGCT CTTGACCTCA
13701 AGTGATATGT CTGCCTCAGN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
13751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
13801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
13851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
13901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
13951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

FIGURE 3, page 4 of 22


```

17751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
20001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
20051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
20101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
20151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNCCAAATC
20201 AACCAGTTGC ATAAATCACT CCTCTATCTT CCTTGGGGTG GAAAGTGGAT
20251 GGGAGTTATA ATTTGAGTTC TCTTTTGCTT TAGTCCATTG AAGCTGCTAT
20301 TACAAATATC CATAAATGCG GTGGCTTATA AACAGCAGAA ATGAGGCCGG
20351 GTGCGGTGGC TCATGCCTAT AATTCCAGCA CTTTGGGAGG CCAAGGCAGG
20401 TGGATCACCT GAGATCAGTA GTTCAAGACT AGCCTGACCA ACATGGTGAA
20451 ACCCTGTCTC TACTAAAAAT ACAAAAATTT AGCTGGGGGT GGTGGCGGGC
20501 ACCTGTAATC CCAGCTACTC AGGAGGCTGA GGCAGGAGAA TCGCTTGAAC
20551 CCAGGAGGCG GAGGTTGCCG TGAGCTGAGA TCACGCCATT GCATTTCAGC
20601 CTGGGCACAA AGAGTGAAC TCCATCTCAA AATGAAATAA AATAACAGAA
20651 ATGTATTTCT TTACAGTTCT GGAGGTTGGG TGGGCAGTCC CAGATCAGGA
20701 CACTGACAGA TTACGTGTCT GATGGGGGCC CACTTCTGG TGTACCTGC
20751 TGGCTGTGTT CTCACATGGT GGAAGGAACA TGGCAACTTT CTGGGGCCTT
20801 GTTTTTTAAT TTAAAAAAA AAAATATTTT CCTGGCCCTT GCCTGCTGAA
20851 GGAACCTCTT TTATAATGGT ACTTAAAAAT TTTTTTTTTT GAGATGGGGG
20901 TCTCACTCTG TCACCCACGC TGAGTGCAGT ATCACAATCT GAGCTCACTG
20951 CAACCTCTGC CTCCCTGGCT TAAGCGATCC TCCACCTCA GCCTCCTGAG
21001 TACGTGTGAC CATAGGCCCA TGGCACAAAG CCCAGCTAAT TTTTGTATT
21051 TTTAGTAGAA ATGTGGTTTC ACCATGTTGC ATAGGCTGGT CTCGAATTC
21101 TGAAGTCAAG TGATCTGCCT GCCTTGGCCT CCCAAAGTGC TGGGATTCTA
21151 GGTATGAGCC ACCCTGCTCG GCCTATAATG GCACTTTCCT ATCCCATTGA
21201 TGAGGCTCTA CTCTCATGAC CTAATCATCT CCCAAAGGCC CTAAGGCCTC
21251 CTGATACCAT CACCTTTGGG GTTAGGTTTT AACATATACA TTTTGGGGGG

```

FIGURE 3, page 6 of 22

21301 ACACAGACAT TTTAGACCAT AGCACCTCCA TTGAAAGGAA ACATTTCTGA
21351 CACCTGGCTA TCTCAAAGGG CCCTTTCAGT TCCCCTGCAG GCTGCATTCC
21401 CACATCACCA ACAAGAGCAG CGACACTCAC TCAGAGGTTA AATAACTTGT
21451 CCAGAGTCAC AGCAGTAATG AATGACAGAG CTGGGGCTTG AATCCAGGCG
21501 TCCTCCTAGA GCCTGGATTG TGTGTAGTGA GTGAAAGCTG ACTCCTGGGA
21551 GACTTCTGCG TGGTCTGGT TCTCTCTCCA GACTGCACTG CGCAAGTTTC
21601 TCTTCCGTAT GGTCCCTAGG GTATTACAAA GACAGTGCCG CTGCCTGTCA
21651 GGTGTTTTTA TTACCAGATG AGGTCAATGGC CTCAGGAACC CTGTAGGAAG
21701 CTGAGTTCAG AGTCTTTGAG CAGGCTTTAG GGAGGTTCCA GCTTCCCACC
21751 ACCAAGCCCC AGGTGGATTG TTACAGACTC TAGCCTCAGG GTGGGGGGTC
21801 TGGGAAGATGA GGTTCGCGGG TCGGATATTG TGCCCAATTG GCCCTCCTT
21851 GCTCAATCTG TTTCTGCAGG TATTGCTGAC TACAGACCCA AGGATGGAGA
21901 AACCATTGAG CTGAGGCTGG TTAGCTGGTA GCCCTGAGC TCCCTCATCC
21951 CAGCAGCCTC GCACACTCCC TAGGCTTCTA CCCTCCCTCC TGATGTCCCT
22001 GGAACAGGAA CTCGCCTGAC CCTGCTGCCA CCTCCTGTGC ACTTTGAGCA
22051 ATGCCCCCTG GGATCACCCC AGCCACAAGC CCTTCGAGGG CCCTATACCA
22101 TGGCCCACTT TGGAGCAGAG AGCCAAGCAT CTTCCTGGG AAGTCTTTCT
22151 GGCCAAGTCT GGCCAGCCTG GCCCTGCAGG TCTCCCATGA AGGCCACCCC
22201 ATGGTCTGAT GGGCATGAAG CATCTCAGAC TCCTTGCCAA AAAACGGAGT
22251 CCGCAGGCCG CAGGTGTTGT GAAGACCACT CGTTCTGTGG TTGGGGTCTT
22301 GCAAGAAGGC CTCCTCAGCC CGGGGGCTAT GGCCCTGACC CCAGCTCTCC
22351 ACTCTGCTGT TAGAGTGGCA GCTCCGAGCT GGTGTGGCA CAGTAGCTGG
22401 GGAGACCTCA GCAGGGCTGC TCAGTGCCTG CCTCTGACAA AATTAAAGCA
22451 TTGATGGCCT GTGGACCTGC TACAGTGGCC TGGTGCCTCA TACTCTCAG
22501 GTGCAGGGGC AGGGACAAGA GAAGGGGGAA GTAACCCCAT CAGGGAGGAG
22551 TGGAGGGTGC CTGAGCCGCC ATGTGGGCAT TGGGGGAGTG ATGGGAATGC
22601 CAGCAGTGAT GACGTTGACT ACTGACTGAG CACCCACTAC TATGACTGAG
22651 CACTCACTCG CTAGATACTA TCTTGAACAT CTCTGTGAGG TTGTTGATAT
22701 TTTCAATTTT ATCTGTGCTT TACAAATCAG GAAACTGGGA GGCCGGGGCGT
22751 GGTGGCTCAC GCCTGTAATC CCAGCACTTT AGGAGGCCAA GGCAGGTGGA
22801 TCACAAGGTC AGGAGTTTGA GATCAGCCTG GCCAACATGG TGAAACTCCA
22851 TCTTTACTAA AAATACAAAA AATTAGCCAG GCATGGTGTG GCATGCCTGC
22901 ATGCCTGTAA TCCAGTTTAC TTGGGAAGCT GAGGCAGGAG AATTGCTTGA
22951 ACCCTGGAGG CGGAGGTTGT AGTGAGCCGA GATCACGCCA TTGCACTCCA
23001 GCTTGGGCAG GAAGAGAAAC ACTCTCAAAA AAAAAAATAA ATCAGGAAAC
23051 TGGTGCTCAA AAAGGAAAAA TGACTCACCA AGGTCACAGA CTAGGCAGTG
23101 ATGCTGGGGG AACCTGGCTC AGGGGACACA GACCTGGCCT GGGGCAGCCT
23151 TGCAGCTCCT CCACTAAAAA ACTGAAAATG AGGGGCTTCG ATGATGGTTA
23201 TAATCGTATG GCAGAGCCCC AACTCACTG GAGCCCTGGG ACCCAGAAGC
23251 TAGGGTCTCA CTCCTGCTT TTCCACAAGG CACCATTAGG GCATCACCCC
23301 AGGCCTCGGC AGCCACGACG CAGGGATCCT GCCTCTCAT TGGTGGGGG
23351 TTAGGGGCTC TGGGCTGCCC TCTTGAAGAG GGGGTTGAGC CCAGCGAGGC
23401 ACCCCCTATG CTGCACCCCA CCAAGGTTAG GAAGAGGTCC TGTCTCAGT
23451 GGGGCCCTCT GATGAACAGC CCATCAGGTC TGCGTCCACA TGCTTGGAA
23501 GAGATGGTGA CATACTCAA GTCTTGAAG CCGCATATTA AACCACTAG
23551 AGCACCATCT TCAAAACATTT AGGGTCTGAG AAGATAGGGG AAGTAAGCAA
23601 TTTAAACAT TTTCTTATAT TGGGCCAGST GCAATGGCTC ACGTCTGTAA
23651 TCCAGCGCTT TTGGGAGGAC GAGGATCACC TGAGGTGAGG AGTTCAAGAT
23701 CAGCCTGGCC AACATGGAGA AACCCCATCT CTAATAAAA TACAAAAATT
23751 AGCTCAGGCG TGGTGATGTG CACCTGTAAT CCTAGCTATT CAGGAGGCTG
23801 AGGCACAAGA ATTGCTTGAG TCAATATTGC ACCACTGCAC TCCAGCCTGG
23851 GCAACAGCGA GACTCTTGTC TCAAAAAAAA AAAAAGATAT TTGCTGAAAA
23901 GACCCAGCCT GCCAAACTCA GGGGCAGCCA AGGGAGGTAG TGAATGGAA
23951 GTTGAGGCTC AGCGCTCCCA CACCTCCACT GCCCTCAGGC CTCTCTGCCC
24001 TCTTTCCCAT CAGTCAGCTG CTTCTGGGCA TGGTCTGCG AGAGACTTGG
24051 CCTCCTTCCA GTTCAAGCTC CCTCTTAGAT TGTGTCCAC GCCACTGAGT
24101 CTTTGGGACA CTGGGTGAGA TGTCTAGTCT GGCACAATTG GCAGGAATCC
24151 CAAGAAACAG TGTGAGTGAG GGGACAGTCG TGTGAGTGC CCTCCATCTG
24201 GGACTGGGAG GCAGGTCTAT GTCAGGCCTG CATTTAGATC TCTAATGGCT
24251 CCAGACAAGC CCCTTCAGCT CACTAAGCCT GTTTCCTAAC ACAGCTGTGG
24301 GATGGTGTCT TGGTTTACAT AGCACGCGAT ACCATCATAG ATCACATGGG
24351 GAAACTGAGG CCCAGGAGT GATCTGCTGG CACATGCAGT GACAAGAGGA
24401 GAGGCCCATC TCAGCCTTGC AGCAAGGTTG CCAGAAATCG ATTCTCGCCC
24451 CCATCCCGTA AAGATAGCTG GGATTACAGG TGTGCACCAC CATGCCAGC
24501 CTAATTTTTG TATTATTAGT AGAGATGGGG TTTCAACATG TTGTCCAGGC
24551 TGGTCATGAA CTCCTGACCT CAAGTGATCC ACCCGCTTTG GCCTCCCAAA
24601 GTGCTGGGAT TACAAGCATG AGCCACAGTG CCTGGCCTGA CCCTGCTCTT
24651 TTGAAAGACC ATTCCCCCAA ATCTGTGCA CCTGTGTGCC TTTCTTCTCT
24701 CTGCCTCCTC TCAGCTCTGC CCCGCTCTCC TCCCTTCTCC TCTGGCAAA
24751 CCCACTCATC TCTTGAAGCC CTTCTTCCAG GGAAGCCCT GATCATGCTG
24801 CTTTCTCCTG TGGGAGGGAT GAAGGACGTG GCCCACGGAG TTTGTTTTGT

FIGURE 3, page 7 of 22

```

24851 TTTGTTTTGA GATGGAGTTT TGCTCATGTT GCCCAGGCTG GGGTACAATG
24901 GTACGATCTC AGCTCACTGC AACCTCTACG TCCCGGGTTC AAGCGGTTCT
24951 CCTGCCCTTAG CCTCCCCAGT AGCTGGGATT ACTGGCATGA ACCACCACAC
25001 CTGGCTAATT TTGTGTTTTT AGTAGAGATG GGGTTTCTTC ATGTTGGTCA
25051 GGCTGGTCTC GAACTCCCAA CCTCAGGTGA TCTGCCTGCC TCGGCCTCCC
25101 AAAGTACTGG GATTACAGGG TTGAGCCACT GTGCCTGGCC CAGGCCACCG
25151 GAGTTTTAAG AGGCTTCCTG TGGCAGTGGC ATCCAGACGG AGTGCAGAAA
25201 CTCAAAGTTG AAGGCCAGAA GCTCAGGGAA GGGGGAGTGT GAGTTGAGGA
25251 GTCTCTTGGC TGCCAGGGCC AGAAACCGAA CTCCAAGCCT CTCCACAACA
25301 GCGGGTGTAG AGCATGTAGA ATCAGAGAGG AGGCTGAGCC ATGCAGCCCC
25351 GAGAAGAGGG GAATGCCACT GAGCCACAGA GACCCAGTGC CACTGCCAGG
25401 TGTCTCTGCC TCCACTTCCC ATGACCCGGC CTGTCTCTGT ATGCAGGCTT
25451 CACCCTCTCT CGTTGTACAT TGTACACATT CTAGGTGACA CCAGCAGCTT
25501 CTGATTCTCA TCTCCCATAA CATCAGCCCC CCAGAGAGGG GACAACTGCT
25551 GAGCTGTATA CATAATAGAT GCCCCTTTCC TGGAGGCCAT GGTATGGTC
25601 AGCGTGGAGA GGATGAAGCC TGAGCAGGCA GGATCGGGGG TCTAGAGGGG
25651 AAGGAGGTGG AAGTTGAGAT CACAGACCTG TGGTCAGGTG GCCTGGGAAG
25701 GGTGACGA GTGTCGGCCC AAAGAGCTTG GAAGGGATTG TGCTGCTGTG
25751 GGTGAGCACT GCCTCTCCCC TTAGGGACAA CAGCCACCTC TTCTCTCCCC
25801 ATTTGCCCTT CCCTTCTGTA GATATGAAAC ACAGGCCTCC TTGTCAGGCC
25851 CCTACTTAAC TCCCGTGATG GGGAAAGCGG CCGGAGAAAG GGAGTTCTGG
25901 CAGCTTCTCC GAGACCCCAA CACCCCACTG TTGCAAGGTG AGTCATGGCC
25951 TGACACTCTG GATGTGTCCC CTACCCCAAG CTTACTCAGC CAAGAGGCTT
26001 CATCAACTCA CCCCAGCTTT CCCTAGCACC CTCCTGGGCC ACACCTTCAC
26051 AAAATCACTG ATGCTCAAAG TTGGATATAA TATATTGAAC TGAAGCCTTA
26101 GCATTTTAT GCAAGTACT GTGGAAATTC TAGGAAACCA GACAGATTAC
26151 AAAAAAAAAA AAAAATAAGA AGAAAATTAA CATCACCTAG GATATACTAC
26201 CTAGGAATAA CGTCTTTTAT TTTGAGATGG AGTTTCGCTC TTGTTGCCCA
26251 GGCTGGAGTG CAGCGGTATG ATCTCGGCTC GCTGCAACCT CCGCCTCCTG
26301 GGTTCATGTG ATTCTTCCAC CTCGGCCTTC CTAGAGCCCA AGTGGTCTGC
26351 CTGCCTCTGC CTCCCAAAGT TCTGGGATTA CAGGCATGAG CCACCGCACC
26401 CAGCCAAAAT TACTTAACTT TTCTTCTAGA TACTTTTTAA AAATATGGCA
26451 GTAAGTTTTT CATAAAAAAT GGAGCCATGC TATCCAGTGG AAATTTAATG
26501 TTGCCACAT GTATAACTTA AAAATTTCAT ATATGTGTAT ACATATATAT
26551 GAAATATATA TATACAGACA CACATATATA TGTATACATA TATATACACA
26601 TATATATGTA TACATATATA CACACATATA TGTATACATA TATATACACA
26651 CATATACACA TATATACACA CACATACATA TATACACACA CATATATACA
26701 CACATATATA CACACATGCA CACATATATA TGTATACATA TATACACACA
26751 TGTATACGTA TATATACACA CATATATACA CACATATATA TACACACATA
26801 TACACACATA CACACACATA TATACACACA TATACACACA CATATATACA
26851 CACATATATA TGTATACATA TATATACACA CATATATACA CATACACACA
26901 TACATATATA CACATATACA CATATACACA CACATATACA CACATGTATA
26951 CATATATATA CACACATGTA TACATATGTA TACACACACA TATATGTATA
27001 CATATATACA CACATACATA TGTGTACATA TATACACACA TACATATGTA
27051 TACATATATA CACACAT

```

(SEQ ID NO:5)

Isoform 1:

FEATURES:

Exon: 2031-2094
 Intron: 2095-5569
 Exon: 5570-5762
 Intron: 5763-7571
 Exon: 7572-7741
 Intron: 7742-10000
 Exon: 10001-10173
 Intron: 10174-10298
 Exon: 10299-10485
 Intron: 10486-12027
 Exon: 12028-12193
 Intron: 12194-25821
 Exon: 25822-25939

Allelic Variants (SNPs):

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
921	C	T	Beyond ORF(5')			
1781	C	T	Beyond ORF(5')			
1850	G	A	Beyond ORF(5')			
2839	A	G	Intron			
3730	G	A	Intron			
6631	G	A	Intron			
6945	-	A	Intron			
6952	A	T	Intron			
7457	G	A	Intron			
7830	T	A	Intron			
8089	T	C	Intron			
8551	C	T	Intron			
9269	G	C	Intron			
9362	C	T	Intron			
9782	G	T	Intron			
11493	G	A T	Intron			
12260	A	G	Intron			
13086	T	C	Intron			
13183	T	C	Intron			
21240	C	G	Intron			
21695	A	G	Intron			
22058	C	T	Intron			
22233	C	A	Intron			
22245	C	-	Intron			
22375	C	T	Intron			
23042	A	- T	Intron			
23344	T	C	Intron			
23873	A	-	Intron			
24764	G	T	Intron			
24939	T	C	Intron			
24945	G	A	Intron			
25092	C	T	Intron			
25428	T	G	Intron			
25513	C	T	Intron			
25684	C	T	Intron			
26165	A	-	Beyond ORF(3')			

Context:

DNA

Position

921 TTGGAGATATTTTAAGGTCATAGTGTCTTCACAAATTGAGCTGAAAGGGAAGTGTAGGA
 TGATCTTGCCTAACCCCTCTCATCTCACACAGGAAGAAGTATTTTAACTCGAGAGGTTAA
 GTGACCTGGCCAAAGTCACACAGCCACCACTAGTTAACTCGTATACATTGATTCTCCTGT
 GGGGCTGGGCAGATGAGGAATCTTTTGTCTCTCCCTGTTGCAGAGATTTTTTTGAG
 GTTACTTTCCGAGTTCTGGCAAGTACCCCTGCTTCTGGTAGCTTTGTGTCTCGATTCAAT
 [C,T]
 TCATTCTTTTATTTTATTTTATTTTGTGAGACAGGGTCTCACTTTGTACCCCAAGCTGGA
 GTGCAGTGGTGTAATCTTGGCTCACTGTAGCCTCCACCTCTTGGGTTCAAGCGATCCTCC

TGCTTCAGCCCCCAAGTAGCTGGGATTACAGACGTCTGCCACCACGCCAGGCTAATTTA
TGGTTTTTGTATGTGTTTTTGTGTTTTTGTAGAGACAGTGTTCCTCCCATGTGCCCAG
GCTGGTCTCCAACCTCTGAGCTCAAGTGATCTGCCCGCTCAGCCTTTCAAAGTGCTAGG

1781 ACAGCTGACTCCAGCAATGCTGCTCACGTGACCACTGCAGCTGCAGCTCCCGTTCCACTC
CTTGTCTTGGCTAGGTGGGCACTACCAGGGGCTCCTTTGGTAAGGAGTACCGGGTAGGC
ACCCGGTCTCTGCAATCCACCCTGGAACAGCTGGGGGGACAGCAGACAGGCACGGTCGG
ACAGACTTGACAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCGGGCTCTTTAAGCAGGA
ACGTGAATGGCTCAAGATGTCTCACATGGTCCCCTAGCCCTCCTCCTCCCTTTGTTC
[C, T]
TACCTCCAGGAGGGTCTCTGCCCTTCTCTCTGTCTTTGGCCTTATGTTCCTCCG
CACCACAGGCCTTCCCCCGCCCCACCCCTCTGCAGACTTAGCCGTGCATTGCAGGCATGG
AGGATTAATCAGTGACAGGAAGCTGCGCTCTCTCGGAGCGGTGACCAGCTGTGGTCAGGAG
AGCCTCAGCAGGGCCAGCCCCAGGAGTCTTTCCCGATTCTTGTCTACTGCTCACCACCT
GCTGCTGCCATGAGGCACCTTGGGGCCTTCTCTCTCTTCTGGGGTCTTGGGGGCCCTC

1850 GGCTAGGTGGGCACTACCAGGGGCTCCTTTGGTAAGGAGTACCGGGTAGGCACCCGGTCC
TGCCAATCCACCCTGGAACAGCTGGGGGGACAGCAGACAGGCACGGTCGGACAGACTTG
ACAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCGGGCTCTTTAAGCAGGAACGTGAATG
GCCTCAAGATGTCTCACATGGTCCCCTAGCCCTCCTCCTCCCTTTGTTCCTACCTCCA
GGAGGCTGTCTGCCCTTCTCTCTGTCTTTGGCCTTATGTTCCTCCGCCACCACAG
[G, A]
CCTTCCCCCGCCCCACCCCTCTGCAGACTTAGCCGTGCATTGCAGGCATGGAGGATTAAT
CAGTGACAGGAAGCTGCGTCTCTCGGAGCGGTGACCAGCTGTGGTCAGGAGAGCCTCAGC
AGGGCCAGCCCCAGGAGTCTTTCCCGATTCTTGTCTACTGCTCACCACCTGCTGCTGCC
ATGAGGCACCTTGGGGCCTTCTCTCTTCTGGGGTCTTGGGGGCCCTCCTGAGATG
TGTGTGAGTAACTCGCCTCTATCTGTGCCTCTTCTCTCTGGGTCTTAGTGGGGTGG

2839 AACATAGGGAGACCCCATCTCTACAAAAATAAAAAAATTAAAAATAGCTGGGCATGG
GGAAGACTTTCTGAAGACCAAGAGGACACATGGGAGCTGAAACTCGAAGGAAGAAAAGGA
GCTGGCAGGAAGAGGAGTGGGGGACACACATTCTAGGCAGCAGGAAGTGAGCCTTCGGAGG
TCCTGCCTGCTCCAGCTCTGTGCCCCAAGGGTCTCTTGGAGCACAGTCTCCTGGGACCT
GTCTATGAGTCTGAGCTTAGAGGCTCAGGGCTGCTCCTTCAGACAGGAGGCAGAAGGCAG
[A, G]
CTTTGGAACTTTGGGCCGCCACGCGCCTTTTCTCCTCCTCTGCACCTAGGATTACGTT
GAGCAATACACTTTCACCCCATGGTCTCTTGGAGCCCTGGGGAAACCCCTGAGAGGTGGG
TGCAGTCAATGTCAGGTGTCAAGTGAAGAAGTCGAGGGTTGGAGGGGCTGAGTGACCCAC
TCAGGGTGTCTCCACCTTTTCCAGAGCTTGTCTGAACCTAGTTTTGAAGCTTGAAGCCTC
GTTTGTTCGTTTTGTTTTTTTGTGAGAGAGGTTCTCCCTCTGTGCCCAGGCTGGAGT

3730 GACACCTCAGGTCTGGGCCCAGGAACCCAGCTCTTGGTTCATGTCCGGACAGTCCCCAG
GGGAGTTCTGGGTTCACCCAGCAAGAGCTCTTCTCCTGGCTGATCTGGTCTCAGCCTT
GGACAGTTAGTCCATTAACTGACCCACAGGAGCCCCAATCCCTTGGGGTCTGGGGAAT
CTTGAAGTGGGGTTTGGGGTGCAAAATATCTGCACTGAGTCACTTAATTGCAACCCAGCCTC
ATTCTTTATCTGTAAGTGGGCTAAGAATGCTCCCTGCCTTCTCCTCGGTGTAGTAC
[G, A]
AGGAAGGATCCCATGACACCTGCTCTCCAGTTTAAAGCTCTATATGTATGTTGTGAAAT
TGACAGGGATCGCTGCACAAACGCTAATGCAAAGTGGGCTCCTGTGCTTCTCTTCTCTT
TCTTCTTCTTTTTTTTTTTTTTTTAAATTTCTTCTAGAGATGAGGTCTCACTATATTGCCCA
GGGTGTTGTTTCAAACTCCTAGGTCAAGCGATCCTCCACCTTGGCCTCCCAAACTGCTG
GTATTACAGGCGTGAGCCACTCTGTCTGGCTCCTATGCTTGTGAATGTCAACAGCAATCA

6631 TGAATAGCTGGGATTACAGGCGTGTGCCACCATGCCAGCTGATTTTTGTATTTTTTGGTA
GAGATAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACCTCTGACCTCAGGTGATCCG
CCTGTCTTGGCTTCCCAAAGTGTGGGATTATAGGCATGAGCCACTGCACCAATCCAAAA
GCAGCATCTTTGTGCTCCCTTTTCAAGAGGCATCACAGAGAGGCTGTTTTGGGGTTTGA
ATGAGAGGCGAAGAATCAGCCATGGAGTGCCTCTTCTCAGACTCCCTCTTGAAGTGG
[G, A]
TGCAGGGGTGGAGAGAAAAGAACTAGGCATAGTGGCTCATACCTGTAATCCCAACATT
TTGGGAGGCTGAGGCAGGAAGATTGCTTGAAGCTCAGGAGTTTGAAGCAGCCTAGGCAAC
ATAGTGAGACCACATCTCTTAAAAAAAAGAAAAAGAAAAAATGAGCCAGGTGTAGTGA
CTCATGCTGTGGTCCCCACTTCTCCGAGGCAAGGTGGGAGGATCTTTTGAAGCTGAG
AAATCAGGGCTACAGTGAAGCATGGTGGCACCCTGCACTCCAGCCTGGGAGACAGAGAG

6945 AGAAAAGAAGACTAGGCATAGTGGCTCATACCTGTAATCCCAACATTTTGGGAGGCTGAG
GCAGGAAGATTGCTTGAAGCTCAGGAGTTTGAAGCAGCCTAGGCAACATAGTGAGACCAC
ATCTCTTAAAAAAAAGAAAAAGAAAAAATGAGCCAGGTGTAGTGACTCATGCTGTGG
TCCCCACTTCTCCGAGGCAAGGTGGGAGGATCTTTGAGGCTGAGAAATCAGGCTAC
AGTGAGCCATGGTGGCACCCTGCACTCCAGCCTGGGAGACAGAGACCCTATCTCAGT
[-, A]
AAAAAAAATAAAAAATATGGCTGGGTGTGGTGGCTCACGCCTGTAATCCCAAGCACTTT

FIGURE 3, page 10 of 22

Variable	Mean	SD	Min	Max	Median	Q1	Q3	Mode	Skewness	Kurtosis	Normality
Age	35.2	12.5	20	65	30	25	35	30	0.5	3.0	Normal
Gender	1.2	0.4	1	2	1	1	1	1	0.2	1.5	Normal
Marital Status	1.5	0.5	1	3	1	1	1	1	0.3	1.8	Normal
Education	12.5	2.0	9	16	12	11	13	12	0.4	2.5	Normal
Income	3000	1500	1000	6000	2500	2000	3000	2500	0.6	3.5	Normal
Occupation	1.8	0.6	1	3	1	1	1	1	0.2	1.5	Normal
Health Status	1.2	0.4	1	2	1	1	1	1	0.2	1.5	Normal
Stress Level	2.5	1.0	1	4	2	1	3	2	0.5	3.0	Normal
Life Satisfaction	3.5	1.5	1	5	3	2	4	3	0.4	2.5	Normal
Resilience	2.8	1.2	1	4	2	1	3	2	0.5	3.0	Normal
Emotional Stability	3.2	1.0	1	4	3	2	4	3	0.4	2.5	Normal
Physical Health	3.8	1.2	1	5	3	2	4	3	0.4	2.5	Normal
Mental Health	3.0	1.0	1	4	2	1	3	2	0.5	3.0	Normal
Overall Well-being	3.5	1.5	1	5	3	2	4	3	0.4	2.5	Normal

9269 AGGCCAGGGTCCCAGGTGCTGGCGGGGTGGCTGCTGGGTGGGGGCAGAGAGGCCAACCCC
TCTGTTTTTTTCCCTCTCAGGGCATGATCACAAGGGCCACCCCCACACTAGCTACTACCA
GTATGGCCTGGGCATTCTGGCCCTGTGTCTCCACCAGAAGCGGGTCCATGACACGCGTGGT
GGACAAACTTCTGTATGCTGTGGAACCTTTCACACAGGCGCCATTCTGTGGGTGAGTA
GGTCAGACCGCTGCCAAGGCCAGGCTGGCACTCCCTCAGTCCCCAGGTCTGCACTGATGAC
[G, C]

TCCATACCCCTGGCCCCACACTCACCTTTTCCTTGGGGCTCCTCCGAATCAAGTCCTTTAG
GGACGAATTGGCGAGGGCTCATGGGTGATGCTCCAGCTGTGAGCCAGCTTTGGAGCTGGT
AGGTGGATCTCTTGGAGCCAGGAGTTCAAGACAACGTGGTGAAACCCCATCTCTACTAAA
AATAAAAAAGTTAGCCGGGCATGGTGGCACATGCCTGTAGTCCAGCTACTCGGGAGGCT
GAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGCTGCAGTGAGTGAGATCGACCA

9362 GGGCCACCCCCACACTAGCTACTACCAGTATGGCCTGGGCATTCTGGCCCTGTGTCTCCA
CCAGAAGCGGGTCCATGACAGCGTGGTGGACAACTTCTGTATGCTGTGGAACCTTTCCA
CCAGGGCCACCATTTCTGTGGGTGAGTAGGTGAGACCGTGCCAGGGCCAGGCTGGCACTCC
CTCAGTCCCCAGGTCTGCATGATGACGTCCATACCCTGGCCCCACACTCACCTTTTCCT
TGGGGCTCCTCCGAATCAAGTCTTTAGGGACGAATTGGCGAGGGCTCATGGGTGATGCT
[C, T]
CAGCTGTGAGCCAGCTTTGGAGCTGGTAGGTGGATCTCTTGGAGCCAGGAGTTCAAGACA
ACGTGGTGAAACCCCATCTCTACTAAAAATAAAAAAGTTAGCCGGGCATGGTGGCACATG
CCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGG
AGGCTGCAGTGAGTGAGATCGCACCACTGCCCTCCAGCCTGGGCAACAGAGTGAGTGAG
ACTCTGTCTCAAAAAATAAAAAATAAAATAAACTCCCTAGTGATTCCAATGTGCAGCT

9782 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCG
GAGGCTGCAGTGAGTGAGATCGCACCACTGCCCTCCAGCCTGGGCAACAGAGTGAGTGA
GACTCTGTCTCAAAAAATAAAAAATAAACTCCCTAGTGATTCCAATGTGCAGC
TAAGTTTGGAAATAGGTGGTATGGGGTCAAGTCTCTTGGGCCTCCCTCCTCCAGTCTCT
CTCCCTAACCTCTAGCCCTCAAGTTGCAGAGTGATCAGCCAACCAAGTTTGGCCAGAAAT
[G, T]
AGCAGTTTCTGGGACACAGGATTTTCAGAGTCCAGACAAGGAAAGTCTTGGGCAGACCA
GGTTGAGTTGGTGCCCTTAGCTGATCTGACCATGTTGCCCTTCTTCTCCAGCCCTCCTG
TGGTTGTCCATAGCTACAAGGGCTGACCCCTCAAGCCCTGCCTGTCTGGCCCCCTTGG
CTCTCCAGCTCATTCATGTTCTGTCTCCCCACTTCAAGACACAGCAGCCATGGCAGGCTT
GGCATTCACTGTCTGAAGCGCTCAAACTTCAACCTGGTGGAGACAACGGATCACCAT

11493 AAAAAAAAAAATGGAGAAGAAGGAGCTGGACATGGTGGCTCGTGCTTATAATCCTAGCA
CTCTGGGAAGCTGAGGCAGATGGATTGCCTGAGCCAGGAGTTTGAGACCAGCCTGGGCA
ACATGGTGAAACCTGTCTTTACTAAAAACGAAAGATTAGCCAGGCATGGTGGTAGACA
CCTATAATCCCAGCTACTAGGGAGGCTGAGGCCACAAGATCACTTGAACCTGGGAGACAG
AGGTTCAGTGAGCCGAGATCGCGCCATTGCACTCCAGCCTGGGCGACAGTGAGACTC
[G, A, T]
GTCTCCAGAAAAAACAAGATGGATAGAGTGAGCCAAGAAGAGGCAGGAAGAACAAGA
CACAGAGGTGCACAGAGTTTGGGGGAATTTTGAGGAATGGTCTTGCAAAAGAGTGGGATC
TGGGAGAATGAGTGGGAGTGGAAGCAGATGAATGAAGAGAAGGTGAGCGCATCAGGGTA
ACAGAGATGCGTTGTGAACAAATGCATGTTCTAGGAAGAGCCCTCTGGAGTGCTAGGTGC
CAGAGAGGTGGGAGGAAGGATACTGGAAGCAGAGAAACCACTGAGGGGCTGATCTTGGG

12260 ACAAGAAGACAAATAATCCAGGCTCTCTGTCTCACACCAGCTGCCCCGCCCTTTCTTCC
TGGCACAGTCACTGTTGGAACCACTGCTGAGACCATTCCTCAGACCCAAGAGATCATCAG
TGTCACGCTGCAGGTGCTTAGTCTCTTGCCGCCGTACAGACAGTCCATCTCTGTTCTGGC
CGGGTCCACCGTGGAAGATGCTCTGAAGAAGGCCCATGAGTTAGGAGGATTACAGTGAGA
TCCCCACCTCCAGTCTCACCACCCCAACCTCACATGCCTGATAACAGGGTCACAGAA
[A, G]
AGACGGGGAACAGAGGAGAGGGTTCCCTCGGGAGAGACACTGGCCCTGCTTCTGCTTCTA
CCTGCTCAGCTCCTTTCTTGCCACGGTGTTATGGAACAGGGAGCCATAGGCCAGCATT
GTCACTGAGAGCAGGCTTTGGAGGCAGAGCCCCCAGTTGGAATCCCAACTTAACCA
GCTAGGTTCCAGGTAGGACCCACAATTACCCGAGGAGAACAGTTGTGCCCTTCCCTGC
AGGGCCAGTGTAAGAGTCCAGGAGTTAGTACACATAGAGATAGTGGCATGTGCTTTTTA

13086 GGCACGTGCCACACAACCTGGGTAATTTTTTTTTTTTTTTTTTTTGGAGATAGGGTCTCTG
TCTGTTGCCAGGCTGGTCTCAAAATCCTGGCCTCAAACCATCCTCACACCTGAGGCGCT
CAAAATATTGGGATTATAGGTGCGAGCCATCATGCTCAGCCAGAATAATAACTGGTTTTT
TTTGTTTTTTTTTTGGAGACAGAGTCTCACTCTATTACCCAGGCTCTGGAGGCCCAACTCG
TGTTTGATTTGTTATTTTATTTATTTATTTATTTATTTTCGAGACAGAGCCTCTCTCTTT
[T, C]
ACCTAGGCTGGAGTGAGTGCGCAATCTCGGCTCACTGCAACCTCCGTCTCCTGGGTTTC
AAGTGATTGTCTGCCTCAGCCTCCTGAGTAGCTGGTGACAGGCGCGTGCCACCATGTC
CCAGCTAATTTTTGTATTTTGTAGTAGACAGGGTTTTACTATGTTGGCCAGCTGGTTTC
TAACTCCTGAACCTCGGGTGATCTGCCTGCCTCGGCTCCCAAAAGTGCTGGGATTACAGGC
ATGGGCTCCGTGCCCGGCCATGTATTTATTTAGGCAAGGTCTCTCTGTTATCCAGGC

13183 ACCATCTCACACCTGAGGCGCTCAAAATATTGGGATTATAGGTGCGAGCCATCATGCTC
AGCCAGAATAATAACTGGTTTTTTTTTTTGTTTTTTTTTTGGAGACAGAGTCTCACTCTATTAC
CCAGGCTCTGGAGGCCCAACTCGTGTGTTGTTATTTGTTATTTTATTTATTTATTTAT
TTCGAGACAGAGCCTCTCTTTCACCTAGGCTGGAGTGAGTGCGCAATCTCGGCTCA
CTGCAACCTCCGTCTCCTGGGTTCAAGTGATTGTCTGCCTCAGCCTCCTGAGTAGCTGG

FIGURE 3, page 12 of 22

[T, C]
GCTACAGGCGCGTGCCACCATGCCAGCTAATTTTGTATTTTGTAGTAGAGACAGGGTTT
TACTATGTTGGCCAGCTGGTTTCTAACTCCTGAACTCGGGTGATCTGCCTGCCTCGGCCT
CCCAAAGTGCTGGGATTACAGGCATGGGCCTCCGTGCCCGGCCATGTATTTTATTTAGGCA
AGGTCTCTCTCTGTTATCCAGGCTGAAGTGCAGTGGCACATTCATAGCTCACTGCAGCCT
CAAAATATCCAAGTAACAGGGACTACAGGCATGCACCACACCCATCTACTTTTTTTT

21240 TCAGCTCACTGCAACCTCTGCCTCCCTGGCTTAAGCGATCCTCCCACCTCAGCCTCCTGA
GTACGTGTGACCATAGGCCCATGGCACAAAGCCCAGCTAATTTTTTGTATTTTGTAGTAGA
AATGTGGTTTCACCATGTTGCATAGGCTGGTCTCGAACTTCTGAACTCAAGTGATCTGCC
TGCCTTGGCCTCCCAAAGTGCTGGGATTCTAGGTATGAGCCACCCTGCTCGGCCTATAAT
GGCACTTTCTATCCCATGATGAGGCTCTACTCTCATGACCTAATCATCTCCCAAAGGC
[C, G]
CTAAGGCCTCTGTATACCATCACCTTTGGGGTTAGGTTTTAACATATACATTTTGGGGGG
ACACAGACATTTTAGACCATAGCACCTCCATTGAAAGGAAACATTTCTGACACCTGGCTA
TCTCAAAGGGCCCTTTTCACTTCCCCTGCAGGCTGCATTCCCACATCACCACAAGAGCAG
CGACTCACTCAGAGGTTAAATAACTTGTCCAGAGTCACAGCAGTAATGAATGACAGAG
CTGGGGCTTGAATCCAGGCGTCTCCTAGAGCCTGGATTCTGTGTAGTGTAGTGAAGCTG

21695 CATTCCCACATCACCACAAGAGCAGCGACACTCACTCAGAGGTTAAATAACTTGTCCAG
AGTCACAGCAGTAATGAATGACAGAGCTGGGGCTTGAATCCAGGCGTCCCTCTAGAGCCT
GGATTCTGTGTAGTGAGTGAAAGCTGACTCCTGGGAGACTTCTGCGTGGTCTCTGGTTCTC
TCTCCAGACTGCACCTGCGCAAGTTTCTTCTCTGATGGTCCCTAGGATTATCAAGACA
GTGGCCCTGCCTGTCAGGTGTTTTTATTACCAGATGAGGTCTAGGCTCAGGAACCCTGT
[A, G]
GGAAGCTGAGTTGAGTCTTTGAGCAGGCTTTAGGGAGGTTCCAGCTTCCCACCACCAA
GCCCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGTGGGGGGTCTGGAAGATGAGGTTG
CGGGGTGCGATATTCTGCCCAATTGCCCCCTCCTTGCTCAATCTGTTTCTGACAGGTATTG
CTGACTACAGACCCAAGGATGGAGAAACCATTGAGCTGAGGCTGGTTAGCTGGTAGCCCC
TGAGCTCCCTCATCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTGATG

22058 CCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGTGGGGGGTCTGGAAGATGAGGTTGCG
GGGTGCGATATTCTGCCCAATTGCCCCCTCCTTGCTCAATCTGTTTCTGCAGGTATTGCT
GACTACAGACCCAAGGATGGAGAAACCATTGAGCTGAGGCTGGTTAGCTGGTAGCCCCCTG
AGCTCCCTCATCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTGATGTC
CCTGGAACAGGAAGTCCGCTGACCTGCTGCCACCTCCTGTGCACTTTGAGCAATGCCCC
[C, T]
TGGGATCACCCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCCACCTTGGAGCAG
AGAGCCCAAGCATCTTCCCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCA
GGTCTCCCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGGC
AAAAAACGGAGTCCGAGGCGCAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTC
CTGCAAGAAGGCTCCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCT

22233 CCTGAGCTCCCTCATCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTG
ATGTCCCTGGAACAGGAAGTCTGCTGACCTGCTGCCACCTCCTGTGCACCTTTGAGCAAT
GCCCCCTGGGATCACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTG
GAGCAGAGAGCCAAGCATCTTCCCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGC
CCTGCAGGTCTCCCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTC
[C, A]
TTGGCAAAAAACGGAGTCCGAGGCGCAGGTGTTGTGAAGACCACTCGTTCTGTGGTTG
GGGTCTTGCAAGAAGGCCCTCCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACT
CTGCTGTTAGAGTGGCAGCTCCGAGCTGGTGTGGCACAGTAGCTGGGAGACCTCAGCA
GGGCTGCTCAGTGCCTGCTGACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTAC
AGTGGCCTGGTGCCTCATACTCCTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGGAAGTA

22245 TCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTGATGTCCCTGGAA
CAGGAAGTCCGCTGACCTGCTGCCACCTCCTGTGCACCTTTGAGCAATGCCCCCTGGGAT
CACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGAGCAGAGAGCC
AAGCATCTTCCCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCAGGTCTC
CCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGCAAAAAA
[C, -]
GGAGTCCGAGGCGCAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTCTGCAAG
AAGGCTCCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCTGTAGAG
TGGCAGCTCCGAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCAGGGCTGCTCAGT
GCCTGCCTCTGACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTACAGTGGCCTGGT
CCTCATACTCCTCAGTGCAGGGGCAGGGACAAGAGAAGGGGGAAGTAACCCCATCAGGG

22375 ACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGAGCAGAGAGCCAAGCATCTTC
CCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCAGGTCTCCCATGAAGGC
CACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGCAAAAAACGGAGTCCGC
AGGCCCGAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTCTGCAAGAAGGCCTCC

FIGURE 3, page 13 of 22

TCAGCCCCGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCTGTTAGAGTGGCAGCTC
[C, T]
GAGCTGGTTGTGGCACAGTAGCTGGGAGACCTCAGCAGGGCTGCTCAGTGCCTGCCTCT
GACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTACAGTGGCCTGGTGCCTCATACTC
CTCAGGTGCAGGGGAGGACAAGAGAAGGGGGAAGTAACCCCATCAGGGAGGAGTGGAG
GGTGCCTGAGCCGCCATGTGGGCATTGGGGAGTGATGGGAATGCCAGCAGTGTGACGT
TGACTACTGACTGAGCACCCTACTATGACTGAGCACTACTCGCTAGATACTATCTTG

23042 GCCGGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTAGGAGGCCAAGGCAGGTGGAT
CACAAAGTCAGGAGTTTGAGATCAGCCTGGCCCAACATGGTGAACTCCATCTTTACTAAA
AATACAAAAAATTAGCCAGGCATGGTGTGTCATGCCTGCATGCCTGTAATCCCAGTTACT
TGGGAAGCTGAGGCAGGAGATTGCTTGAACCTGGAGGCGGAGGTTGTAGTGAGCCGAG
ATCAGCCCATGCACTCCAGCTTGGGCAAGAAGAGAACACTCTCAAAAAAAAAAAAAA
[A, -, T]
CAGGAACTGGTGTCAAAAAGGAAAAGTGACTCACCAAGGTCACAGACTAGGCAGTGAT
GCTGGGGGAACCTGGCTCAGGGGACACAGACCTGGCCTGGGGCAGCCTTGACGCTCCTCC
ACTAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAA
CTCAACTGGAGCCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTTCCACAAGGCA
CCATTAGGGCATCACCCAGGCCCTCGGCAGCCACGACGCAGGGATCCTGCCTCTCATTTGG

23344 AGGAAACTGGTGCTCAAAAAGGAAAAGTGACTCACCAAGGTCACAGACTAGGCAGTGATG
CTGGGGGAACCTGGCTCAGGGGACACAGACCTGGCCTGGGGCAGCCTTGACGCTCCTCCA
CTAAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAAC
TCAACTGGAGCCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTTCCACAAGGCAC
CATTAGGGCATCACCCAGGCCCTCGGCAGCCACGACGCAGGGATCCTGCCTCTCATTTGGT
[T, C]
GGGGGCTTAGGGGCTCTGGGCTGCCCTCTTGAAGAGGGGTTACGCCAGCGAGGCACCC
CCTATGCTGCACCCACCAAGGTTAGGAAGAGGTCCTGTCTCAGTGGGGCCCTCTGATG
AACAGCCCATCAGGTCTGCGTCCACATGCCTTGAAGAGATGGTGACATACTCAAAGTCC
TTGAAGCCGCATATTAAACACCTAGAGCACCATCTTCAAACATTTAGGGTCTGAGAAGA
TAGGGGAAGTAAGCAATTAAACATTTCTTTATATTGGGCCAGGTGCAATGGCTCACGT

23873 GGTCTGAGAAGATAGGGGAAGTAAGCAATTTAAACATTTCTTTATATTGGGCCAGGTGC
AATGGCTCACGTCTGTAATCCCAGCGCTTTGGGAGGACGAGGATCAGGTGAGGTGAGGAG
TTCAAATACGCTGGCCCAACATGGAGAAACCCATCTCTACTAAAAATACAAAATTAG
CTCAGGCGTGGTGTGTCACCTGTAATCCTAGCTATTGAGGAGGCTGAGGCACAGAAT
TGCTTGAGTCAATATTGACCACTGCACTCCAGCCTGGGCAACAGCGAGACTCTTGTCTC
[A, -]
AAAAAAAAAAGATATTGCTGAAAAGACCCAGCCTGCCAACTCAGGGGAGCCAAGG
GAGGTAGTGAAATGGAAGTTGGAGCTCAGCGCTCCACACCTCCACTGCCCTCAGGCCTT
CTCTGCCCTCTTTCCCATCAGTCAGCTGCTTCTGGGCATGGTCTGGCAGAGACTTGGCCT
CCTTCCAGTTCAAGCTCCCTCTTAGATTGTGTCCCACGCCACTGAGTCTTTGGGACACTG
GGTCAGATGTCTAGTCTGGCACAATTGGCAGGAATCCCAAGAAACAGTGTGAGTGAGGGG

24764 ATAGCTGGGATTACAGGTGTGCACCACCATGCCAGCCTAATTTTGTATTATTAGTAGA
GATGGGGTTTACCATGTTGTCCAGGCTGGTCATGAACCTCCTGACCTCAAGTGATCCACC
CGCTTTGGCCTCCCAAGTGCTGGGATTACAAGCATGAGCCACAGTGCCTGGCCTGACCC
TGCTCTTTTGAAGACCATTCCCCAAATTCTGTGCACCTGTGTGCCTTTCTTCTCTCTG
CCTCCTCTCAGCTCTGCCCGCTCTCCTCCCTTCTCCTCTGGCAAAATCCCACTCATCTCT
[G, T]
GAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGGATGAAG
GACGTGGCCACCGAGTTTGTGTTTTGTTTTGAGATGGAGTTTGTCTCATGTGCCC
AGGCTGGGGTACAATGTTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGTTCAAGC
GGTTCTCCTGCCTTAGCTCCCAAGTAGCTGGGATTACTGGCATGAACCACCACTGG
CTAATTTTGTGTTTTTAGTAGAGATGGGGTTCTTTCATGTTGGTCAGGCTGGTCTCGAAC

24939 GACCCTGCTCTTTTGAAGACCATTCCCCAAATTTCTGTGCACCTGTGTGCCTTTCTTCT
CTCTGCCTCCTCTCAGCTCTGCCCCGCTCTCCTCCCTTCTCCTCTGGCAAAATCCCACTCA
TCTCTTGAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGG
ATGAAGGACGTGGCCACGGAGTTTGTGTTTTGTTTTGAGATGGAGTTTGTCTCATG
TTGCCAGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGT
[T, C]
CAAGCGGTTCTCCTGCCTTAGCTCCCAAGTAGCTGGGATTACTGGCATGAACCACCACA
CCTGGCTAATTTTGTGTTTTTAGTAGAGATGGGGTTTCTTCATGTTGGTCAGGCTGGTCT
CGAACTCCCAACCTCAGGTGATCTGCCTGCCTCGGCCTCCCAAGTACTGGGATTACAGG
GTTGAGCCACTGTGCCTGGCCAGGCCACGGAGTTTAAAGAGGCTTCTGTGGCAGTGG
CATCCAGACGGAGTGCAAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGGAGTG

24945 GCTCTTTTGAAGACCATTCCCCAAATTTCTGTGCACCTGTGTGCCTTTCTTCTCTCTGC
CTCCTCTCAGCTCTGCCCCGCTCTCCTCCCTTCTCCTCTGGCAAATCCCACTCATCTCTT
GAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGGATGAAG

FIGURE 3, page 14 of 22

GACGTGGCCACGGAGTTTGTGTTTGTGTTTGTGAGATGGAGTTTGTCTCATGTTGCC
AGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGTTCAAGC
[G, A]
GTTCTCCTGCCTTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCACCACACCTGGC
TAATTTTGTGTTTTAGTAGAGATGGGGTTTCTTCATGTTGGTCAGGCTGGTCTCGAACT
CCCAACCTCAGGTGATCTGCCTGCCTCGCCTCCCAAAGTACTGGGATTACAGGGTTGAG
CCACTGTGCCTGGCCAGGCCACGGAGTTTAAAGAGGCTTCTGTGGCAGTGGCATCCA
GACGGAGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGAAGGGGAGTGTGAGTT

25092 ATCATGCTGCTTCTCCTGTGGGAGGGATGAAGGACGTGGCCACGGAGTTTGTGTTTGT
TTGTTTGTGAGATGGAGTTTGTCTCATGTTGCCAGGCTGGGGTACAATGGTACGATCTCA
GCTCACTGCAACCTCTACGTCCCGGGTTCAAGCGGTTCTCCTGCCTTAGCCTCCCCAGTA
GCTGGGATTACTGGCATGAACCACCACACCTGGCTAATTTTGTGTTTTAGTAGAGATGG
GGTTTCTTCATGTTGGTCAAGGCTGGTCTCGAACTCCCAACCTCAGGTGATCTGCCTGCCT
[C, T]
GGCCTCCCAAAGTACTGGGATTACAGGGTTGAGCCACTGTGCCTGGCCAGGCCACGGGA
GTTTTAAAGAGGCTTCCTGTGGCAGTGGCATCCAGACGGAGTGCAGAACTCAAAGTTGAA
GGCCAGAAGCTCAGGGAAGGGGAGTGTGAGTTGAGGAGTCTCTTGGCTGCCAGGGCCAG
AAACCGAACTCCAAGCCTCTCCACAACAGCGGTTGAGAGCATGTAGAATCAGAGAGGAG
GCTGAGCCATGCAGCCCCGAGAAGAGGGGAATGCCACTGAGCCACAGAGACCCAGTGCCA

25428 AGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGAGTGTGAGTTGAGGA
GTCTCTTGCTGCCAGGGCCAGAAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAG
AGCATGTAGAATCAGAGAGGAGGCTGAGCCATGCAGCCCCGAGAAGAGGGGAATGCCACT
GAGCCACAGAGACCCAGTCCACTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCC
[T, G]
GCCTGTCTCTGTATGCAGGCTTCACCCTCTCTCGTTGTACATTGTACACATTCTAGGTGA
CACCAGCAGCTTCTGATTCTCATCTCCATAACATCAGCCCCCAGAGAGGGGACAACCTG
CTGAGCTGATAACATAATAGATGCCCCTTTCTTGAGGCCATGTCATGGTCAGCGTGGA
GAGGATGAAGCCTGAGCAGGCAGGATCGGGGGTCTAGAGGGGAAGGAGGTGGAAGTT

25513 GGCCAGAAGCTCAGGGAAGGGGAGTGTGAGTTGAGGAGTCTCTTGGCTGCCAGGGCCAG
AAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAGAGCATGTAGAATCAGAGAGGAG
GCTGAGCCATGCAGCCCCGAGAAGAGGGGAATGCCACTGAGCCACAGAGACCCAGTGCCA
CTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCCGGCCTGTCTCTGTATGCAGGCTTCA
CCCTCTCTCGTTGTACATTGTACACATTCTAGGTGACACCAGCAGCTTCTGATTCTCATC
[C, T]
CCCATAACATCAGCCCCCAGAGAGGGGACAACCTGTGAGCTGATAACATAATAGATGCC
CCTTTCCTGGAGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGAGCAGGCAGGA
TCGGGGGTCTAGAGGGGAAGGAGGTGGAAGTTGAGATCACAGACCTGTGGTCAGGTGGCC
TGGAAGGGTTTGACGAGTGTGCGCCCAAAGAGCTTGGAAGGGATTTTGTCTGTGTGGGT
GAGCACTGCCTCTCCCCTTAGGGACAACAGCCACCTCTTCTCTCCCCATTGCTTTTCCC

25684 CCAGTGCCACTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCCGGCCTGTCTCTGTATG
CAGGCTTCACCCTCTCTCGTTGTACATTGTACACATTCTAGGTGACACCAGCAGCTTCTG
ATTCTCATCTCCATAACATCAGCCCCCAGAGAGGGGACAACCTGTGAGCTGATAACAT
AATAGATGCCCCTTTCTTGAGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGA
GCAGGCAGGATCGGGGGTCTAGAGGGGAAGGAGGTGGAAGTTGAGATCACAGACCTGTGG
[C, T]
CAGGTGGCCTGGGAAGGGTTTGACGAGTGTGCGCCCAAAGAGCTTGGAAGGGATTTTGTCT
GCTGTGGCTGAGCACTGCCTCTCCCCTTAGGGACAACAGCCACCTCTTCTCTCCCCATTT
GCCTTTCCCTTCTGTAGATAGAAACACAGGCCTCCTTGTGAGGCCCTTACTTAACTTCC
GTGATGGGGAAGCGCCGAGAAAGGGAGTTCTGGCAGCTTCTCCGAGACCCCAACACC
CCACTGTTGCAAGGTGAGTCATGGCCTGACACTCTGGATGTGTCCCTACCCCAAGCTTA

26165 GTGATGGGGAAGCGGCCGAGAAAGGGAGTTCTGGCAGCTTCTCCGAGACCCCAACACC
CCACTGTTGCAAGGTGAGTCATGGCCTGACACTCTGGATGTGTCCCCTACCCCAAGCTTA
CTCAGCCAAGAGGCTTCATCAACTCACCCAGCTTCCCTAGCACCTCCTGGGCCACAC
CTTCACAAAATCACTGATGCTCAAAGTTGGATATAATATATTGAACTGAAGCCTTAGCAT
TTTTATGCAAGTTACTGTGGAATTTCTAGGAAACCAGACAGATTACAAAAA
[A, -]
CTAGAAGAAAATTAACATCACCTAGGATATACTACCTAGGAATAACGTCTTTTATTTTGA
GATGGAGTTTCGCTCTTGTGCCCCAGGCTGGAGTGCAGCGGTATGATCTCGGCTCGCTGC
AACCTCCGCCTCCTGGGTTTCATGTGATTCTTCCACCTCGGCCTTCTAGAGCCCAAGTGG
TCTGCCTGCCTCTGCCTCCCAAAGTTCTGGGATTACAGGCATGAGCCACCGCACCCAGCC
AAAATTACTTAACTTTCTTCTAGATACTTTTAAAAATATGGCAGTAAGTTTTTCATAA

Isoform 2:

FEATURES:

Exon: 2132-2195
Intron: 2196-5670
Exon: 5671-5863
Intron: 5864-7672
Exon: 7673-7761
Intron: 7762-9149
Exon: 9150-9302
Intron: 9303-10101
Exon: 10102-10274
Intron: 10275-10399
Exon: 10400-10586
Intron: 10587-12128
Exon: 12129-12294
Intron: 12295-25922
Exon: 25923-26040

Allelic Variants (SNPs):

DNA				Protein		
Position	Major	Minor	Domain	Position	Major	Minor
1022	C	T	Beyond ORF(5')			
1882	C	T	Beyond ORF(5')			
1951	G	A	Beyond ORF(5')			
2940	A	G	Intron			
3831	G	A	Intron			
6732	G	A	Intron			
7558	G	A	Intron			
7931	T	A	Intron			
8190	T	C	Intron			
8652	C	T	Intron			
9370	G	C	Intron			
9463	C	T	Intron			
9883	G	T	Intron			
11594	G	A T	Intron			
12361	A	G	Intron			
13187	T	C	Intron			
13284	T	C	Intron			
21341	C	G	Intron			
21796	A	G	Intron			
22159	C	T	Intron			
22334	C	A	Intron			
22346	C	-	Intron			
22476	C	T	Intron			
23143	A	- T	Intron			
23445	T	C	Intron			
23974	A	-	Intron			
24865	G	T	Intron			
25040	T	C	Intron			
25046	G	A	Intron			
25193	C	T	Intron			
25529	T	G	Intron			
25614	C	T	Intron			
25785	C	T	Intron			
26266	A	-	Beyond ORF(3')			

Context:

DNA
Position

1022 TTGGAGATATTTTAAGGTCATAGTGTCTTCACAAATTGAGCTGAAAGGGAAGTGTAGGA
TGATCTTGCCTAACCTCTCATCTCACACAGGAAGAACTATTTTAACTCGAGAGGTTAA
GTGACCTGGCCAAAGTCACACAGCCACCACTAGTTAACTCGTATACATTGATTCTCCTGT
GGGGCTGGGCAGATGAGGAATCTTTTGTCTCTTCCCTGTTTGCAGAGATTTTTTTGAG
GTTACTTTCCGAGTTCTGGCAAGTACCCCTGCTTCTGGTAGCTTTGTGTCTCGATTCAAT
[C,T]
TCATTCTTTTATTTTATTTTATTTTATTTTGTGAGACAGGGTCTCACTTTGTCAACCAAGCTGGA

GTGCAGTGGTGTAACTCTTGGCTCACTGTAGCCTCCACCTCTTGGGTTCAAGCGATCCTCC
TGCCTCAGCCCCCAAGTAGCTGGGATTACAGACGTCTGCCACCACGCCAGGCTAATTTA
TGGTTTTTTGTATGTGTTTTTTGTGTTTTTGTAGAGACAGTGTTCCTCATGTTGCCAG
GCTGGTCTCCAACCTCTGAGCTCAAGTGATCTGCCCGCTCAGCCTTTCAAAGTGCTAGG

1882 ACAGCTGACTCCAGCAATGTGCTCACGTGACCACTGCAGCTGCAGCTCCCGTTCCACTC
CTTGTCTCTGGGCTAGGTGGGCACTACCAGGGGCTCCTTTGGTAAGGAGTACCGGGTAGGC
ACCCGGTCTCTGCCAATCCCACTGGAACAGCTGGGGGGACAGCAGACAGGCACGGTCGG
ACAGACTTGACAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCGGGCTCTTTAAGCAGGA
ACGTGAATGGCCTCAAGATGTCTCACATGGTCCCCTAGCCCTCCTCCTCCTTTGTTCC
[C, T]
TACCTCCAGGAGGGCTGCTCTGCCCTTCCTTCTCTGTTCTTTGGCCTTATGTTCCCCGC
CACCACAGGCCTTCCCCGCCCCACCCCTCTGCAGACTTAGCCGTGCATTGCAGGCATGG
AGGATTAATCAGTACAGAAAGCTGCGTCTCTCGGAGCGGTGACCAGCTGTGGTCAGGAG
AGCCTCAGCAGGGCCAGCCCCAGGAGTCTTTCCGATTCTTGCTCACTGCTACCCACCT
GCTGCTGCCATGAGGCACCTTGGGGCTTCCTCTTCTCTGTTCTGTTGGGGCTCTGGGGCCCTC

1951 GGCTAGGTGGGCACTACCAGGGGCTCCTTTGGTAAGGAGTACCGGGTAGGCACCCGGTCC
TGCCAATCCCACTGGAACAGCTGGGGGGACAGCAGACAGGCACGGTCGGACAGACTTG
ACAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCGGGCTCTTTAAGCAGGAACGTGAATG
GCCTCAAGATGTCTCACATGGTCCCCTAGCCCTCCTCCTCCTTTGTTCCCTACCTCCA
GGAGGGCTGCTCTGCCCTTCCTTCTCTGTTCTTTGGCCTTATGTTCCCCGCCACCACAG
[G, A]
CCTTCCCCCGCCCCACCCCTCTGCAGACTTAGCCGTGCATTGCAGGCATGGAGGATTAAT
CAGTGACAGGAAGCTGCGTCTCTCGGAGCGGTGACCAGCTGTGGTCAGGAGAGCCTCAGC
AGGGCCAGCCCCAGGAGTCTTTCCCGATTCTTGCTCACTGCTCACCACCTGCTGCTGCC
ATGAGGCACCTTGGGGCTTCCTCTTCTTCTGTTGGGGCTCTGGGGCCCTCACTGAGATG
TGTGGTGAGTAACCTCGCCTCTATCTGTGCTCTTTCTCTGGGTCCTTAGTGGGGTGG

2940 AACATAGGGAGACCCCATCTCTACAAAAAATAAAAAAATAAAAAATAGCTGGGCATGG
GGAAGACTTTCTGAAGACCAAGAGGACACATGGGAGCTGAAACTCGAAGGAAGAAAAGGA
GCTGGCAGGAAAGGAGTGGGGGACACACATTCTAGGCAGCAGGAAGTGAACCTTCGGAGG
TCCTGCTGCTCCAGCTCTGTGCCCAAGGGGTCTCTTGGAGCACAGTCTCCTGGGACCT
GTCTATGAGTCTGAGCTTAGAGGCTCAGGGCTGCTCCTTCAGACAGGAGGCAGAAGGCAG
[A, G]
CTTTGGGAACCTTTGGGCGGCCACGCGCCTTTTCTCCTCCTCTGCACCTAGGATTACGTT
GAGCAATACACTTTACCCCATGGTCTCTTGAGACCTGGGGAAACCTGAGAGGTGGG
TGCAGTCACTGTCAGGTGTCAAGTGAAGAAGTCAAGGGTGGAGGGGCTGAGTGACCCAC
TCAGGTGCTTCCACTTTTCCAGAGCTTTGCTGAACTTAGTTTTTAGAACTGAAGCCTC
GTTTGTCTTCGTTTTGTTTTTGTGTGAGAGAGGTCTCCCTCTGTTGCCAGGCTGGAGT

3831 GACACCTCAGGTCTGGGCCCAGGAACCCAGCTCTTGGTTCATGTCCGGACAGTCCCCAG
GGGAGTTCTGGGTTCAACCAGCAAGAGCTCTTCTCCTGGCTGATCTGGTCTCAGCCTT
GGACAGTTAGTCCATTAACCTGACCCACAGGAGCCCCAATCCTTTGGGGTCTGGGGAAT
CTTGAATGGGGTTTGGGGTGCAAAATATCTGCACTGAGTCACTTAATTGCACCCAGCCTC
ATTCTTTTATCTGTAAAGTGGGCTAAGAATGCTCCCTGCCTTCTCCTCGGTGTAGTAC
[G, A]
AGGAAGGATCCCATGACACCTGCTCTCCAGTTTAAAGCTCTATATGTATGTTGTGAAAT
TGACAGGGATCGCTGCACAAACGCTAATGCAAAGTGGGCTCCTGTGCTTCTTTCTCTT
TCTTCTTCTTTTTTTTTTTTTTAATTTCTTCTAGAGATGAGGTCTCACTATATTGCCCA
GGGTTGGTTTCAAACTCCTAGGGTCAAGCGATCCTCCACCTTGGCCTCCCAAACTGCTG
GTATTACAGGCGTGAGCCACTCTGTCTGGCTCCTATGCTTGTGAATGTCAACAGCAATCA

6732 TGAATAGCTGGGATTACAGGCGTGTGCCACCATGCCAGCTGATTTTTGTATTTTTTGGTA
GAGATAGGGTTTCAACATGTTGGCCAGGCTGGTCTTGAACCTCTGACCTCAGGTGATCCG
CCTGTCTTGGCTTCCCAAAGTGTGGGATTATAGGCATGAGCCACTGCACCAATCCAAAA
GCAGCATCTTTGTGCTCCCTTTTCAAGAGGCATCAGAGAGGCGCTGTTTTGGGGTTTGA
ATGAGAGGGCAAGAATCAGCCATGGAGTGCCTCTTTCTCAGACTCCCTCTTGAGAAGTGG
[G, A]
TGCAGGGGTGGAGAGAAAAGAAGACTAGGCATAGTGGCTCATACCTGTAATCCCAACATT
TTGGGAGGCTGAGGCAGGAAGATTGCTTGAGCTCAGGAGTTTGAGACCAGCCTAGGCAAC
ATAGTGAGACCACATCTCTTAAAAAAGAAAAAGAAAAAATGAGCCAGGTGTAGTGA
CTCATGCCTGTGGTCCCCACTTCTCCGGAGGCAAGGTGGGAGGATCTTTTGGGCTGAG
AAATCGAGGCTACAGTGAGCCATGGTGGCACCCTGCACTCCAGCCTGGGAGACAGAGAG

7558 AAGAATCGCTTGAACCTCGGGAGGCGGAGGTGTCAGTGAGCTGAGAACATGCCACTGCACCT
CCAGCCTGGGCAACAAGAGCGAAACTCTGTCTCAAAGAAAATAAATAAATAAATAAATAA
AATAAAAAAGGAGGGGGCATATGGGTGAAGTATGGACAAAATAGTGGGGCAGGCACAGAT
GATCTGGACACAGGAGCCCTTGGAGTTTATTCTTGAATCTAACTGTTTCATCTTTATTA
TATTTGTGTCATACCTCACAACAACATAGCCAACACACCTCCTTTTGGAGCTTTTATC
[G, A]

FIGURE 3, page 17 of 22

AAGTTTCCCAGTGTAAAGATTTTTTCCCGCTTTGTGATGCGGGTGGGGTGGGTGCTGTAA
 GCAGGCTTACGGGTGGCAGTTTCTCACAAGGCATTAACTGGCCTTGTCTAGGTCTGC
 CTTACAGCGAGGATGACGGTGACTGCCAGGGCAAGCCTTCCATGGGCCAGCTGGCCCTCTA
 CCTGCTCGCTCTCAGAGCCAACGTGTAGTTTGTGAGGGGCCACAAGGGGGACAGGCTGGT
 CTCACAGCTCAAATGGTTCTCGAGGATGAGAAGAGAGCCATTGGTGAGCAGACACCATC

7931 GGTGGCAGTTTCTCACAAGGCATTAACTGGCCTTGTCTAGGTCTGCCTTCAGCGAGGA
 TGACGGTGACTGCCAGGGCAAGCCTTCCATGGGCCAGCTGGCCCTCTACCTGCTCGCTCT
 CAGAGCCAACGTGTAGTTTGTGAGGGGCCACAAGGGGGACAGGCTGGTCTCAGAGCTCAA
 ATGGTTCTCGAGGATGAGAAGAGAGCCATTGGTGAGCAGACACCATCCGCTGGGGGTGG
 GGAGCAGCTGGGAGGGCTCATCAGATGATATTCTCCAATGAGAATCAGAAGTTTGGGTTT
 [T, A]
 CTCCTCCAGGCGTCTTCCCACCATCCATTCTGCCATCTCACTGCCTACGTAGAGGCTCG
 AACCTGTCCCAGATAGCCATCCTTGACCCAGCTTTTCCCGCGCTGCACACATACTATTGAC
 AGGTGTGTTTCGTGGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG
 TCTTGTCTGCCAGGCTGGAGTACAATGGCGCAATCTCAGCTCACCGCAATCTCTGCCTCC
 TGGGTTCAAGCAATCTCTTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCATGCGCC

8190 ATCAGATGATATTCTCCAATGAGAATCAGAAGTTTGGGTTTCTCCCAGGCGTCTTCC
 CACCATCCATTCTGCCATCTCACTGCCTACGTAGAGGCTCGAACCTGTCCCAGATAGCCA
 TCCTTGACCCAGCTTTTCCCGCGCTGCACACATACTATTGACAGGTGTGTTTCGTGGTTT
 TTTGTTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG
 AGTACAATGGCGCAATCTCAGCTCACCGCAATCTCTGCCTCCTGGGTTCAAGCAATTCTC
 [T, C]
 TGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCATGCGCCACCACACCCAGCTAATTTT
 GTATTTTGTAGTAGAGTGGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAACTCCTGACC
 TCAGGTGATCCGCTTGCCTTAGCCTCCGAAAGTGTGGGATTACAGGCATGAGCCACTGC
 GTTAGGCCCACTGACAAGCCTTGTATTGGCTAGCCACCAAGATTGACTTGATTATCCACC
 TTCGGGACAACCTGGACAGCCTGCTTATGACTTACGCCATAGTCTGTCTCTACTAGCTCTC

8652 TACAGGCATGAGCCACTGCGTTAGGCCCACTGACAAGCCTTGTATTGGCTAGCCACCAAG
 ATTGACTTGATTATCCACCTTCGGGACAACCTGGACAGCCTGCTTATGACTTACGCCATAG
 TCTGTCTCTACTAGTCTCTCTGCCCTGACTTGACCCAGCATAACAAGCCAGAGCCAGCC
 TTTTCAATATAAACCTGATCTTGCTGGCACTGCTTAAACCTGTCAGGGGCCCTCGCACTGC
 TCCATGGCCCCAGCCTGTCTACCTTACCTTCTGCCAGGCTCTGCTCATCCATTCTCTGC
 [C, T]
 TCCCACACACCTGCCCTCTGTGGGCTCCAGCCATACCATCTCTCAACTCATAAGCCAGTT
 TTTTCATACAGGCTCCCTCCATCTGACTGGCTTCCCTGCGTGCAGTTCACTCCTGCTCT
 ACCTTTGGCTCTGCTCCACCCATCCTCAGCCGTCTCCAGCATTACCTCCTTGGAGAATC
 CTGCCCTGACTTCCAGCCACCCAAATATCACTACTTGGTCTGCATTCTCGTTGCAATTG
 CAGTCGATGAGCAATTGCTGTGGTTGAGGCCCCGAAGTGCAGCAAGTGCCTGTCTGCCATG

9370 AGGCCAGGGTCCCAGGTGCTGGCGGGGCTGGCTGCTGGGTGGGGGCAGAGAGGCAACCCC
 TCTGTTTTTTTCCCTCTCAGGGCATGATCACAAGGGGCCACCCACACTAGCTACTACCA
 GTATGGCCTGGGCATTCTGGCCCTGTGTCTCCACCAGAAGCGGGTCCATGACAGCGTGGT
 GGACAAACTTCTGTATGCTGTGGAACCTTCCACCAGGGGCCACCATCTGTGGGTGAGTA
 GGTACAGCCTGCCAAGGCCAGGCTGGCACTCCCTCAGTCCCAGGCTCTGCACTGATGAC
 [G, C]
 TCCATACCTTGGCCCCACACTCACCTTTCTTGGGGCTCCTCCGAATCAAGTCCCTTAG
 GGACGAATTGGCGAGGGCTCATGGGTGATGCTCCAGCTGTGAGCCAGCTTTGGAGCTGGT
 AGGTGGATCTCTTGAGGCCAGGAGTTCAAGACAACGTGGTGAACCCCATCTCTACTAAA
 AATAAAAAAGTTAGCCGGCATGGTGGCACATGCCTGTAGTCCAGCTACTCGGGAGGCT
 GAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGCTGCAGTGAGTGAGATCGCACCA

9463 GGGCCACCCCACTAGCTACTACAGTATGGCCTGGGCATTCTGGCCCTGTGTCTCCA
 CCAGAAGCGGGTCCATGACAGCGTGGTGGACAAACTTCTGTATGCTGTGGAACCTTTCCA
 CCAGGGCCACCATCTGTGGGTGAGTAGGTGACACCGTGCCAGGCCAGGCTGGCACTCC
 CTCAGTCCCAGGCTGCACTGATGACGTCCATACCCTGGCCCCCACTCACCTTTCTCT
 TGGGGCTCCTCCGAATCAAGTCTTTAGGGACGAATTGGCGAGGCTCATGGGTGATGCT
 [C, T]
 CAGCTGTGAGCCAGCTTTGGAGCTGGTAGGTGGATCTCTTGAGGCCAGGAGTTCAAGACA
 ACGTGGTGAACCCCATCTCTACTAAAAATAAAAAAGTTAGCCGGGCATGGTGGCACATG
 CCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGG
 AGGTGCACTGAGTGGAGATCGCACCCTGCCCTCCAGCCTGGGCAACAGAGTGAGTGAG
 ACTCTGTCTCAAAAAATAAAAAATAAAATAAACTCCCTAGTGATTCCAATGTGCAGCT

9883 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCG
 GAGGCTGCAGTGAGTGGAGATCGCACCCTGCCCTCCAGCCTGGGCAACAGAGTGAGTGA
 GACTCTGTCTCAAAAAATAAAAAATAAAATAAACTCCCTAGTGATTCCAATGTGCAGC
 TAAGTTTGGAAATAGGTGGTATGGGGTCAAGTCCCTTGGGGCTCCCTCCTCCAGTCCCT
 CTCCTAACCTCTAGCCCTCAAGTTGCAGAGTGATCAGCCAACAGGTTTGGCCAGAAAT

FIGURE 3, page 18 of 22

[G, T]
AGCAGTTTCCTGGGACACAGGATTTTCAGAGTCCAGACAAGGAAAGTCTTGGGCAGACCA
GGTTGAGTTGGTGCCCTTAGCTGATCTGACCATGTTGCCCTTCTTCTCCAAGCCCTCTG
TGGTTGTCCATAGCTACAAGGCCCTGACCTCAAGCCCTGCTGTCTGGCCCTTTGG
CTCTCCAGCTCATTGCATGTTCTGTCCCCCACTTCAAGACACAGCAGCCATGGCAGGCTT
GGCATTACCTGTCTGAAGCGCTCAAACCTTCAACCTGGTCTGGAGACAACGGATCACCAT

11594
AAAAAAAAAATGGAGAAGGAAGCTGGACATGGTGGCTCGTGCTTATAATCCTAGCA
CTCTGGGAAGCTGAGGCAGATGGATTGCCCTGAGCCAGGAGTTTGAGACCAGCTGGGCA
ACATGGTGAAACCTGTCTTTACTAAAATACGAAAGATTAGCCAGGCATGGTGGTAGACA
CCTATAATCCCAGCTACTAGGGAGGCTGAGCCACAAGAATCACTTGAACCTGGGAGACAG
AGGTTGCAGTGAGCCGAGATCGCGCCATTGCACTCCAGCCTGGCGACAGTGTGAGACTC
[G, A, T]
GTCTCCAGAAAAACAAGAATGGATAGAGTGGAGCCAAGAAGAGGCAGGAAGAACAAGA
CACAGAGGTGCACAGAGTTTGGGGGAATTTGAGGAATGGTCTTGCAAAAGAGTGGGATC
TGGGAGAATGAGTGGGAGTGGAAAGCAGATGAATGAAGAGAAGGTGAGCGCATCAGGGTA
ACAGAGATGCCTTGTGAACAAATGCATGTTCTAGGAAGAGCCCTCTGGAGTGCATAGGTGC
CAGAGAGGTGGGAGGAAGGATACTGGAAGCAGAGAAACAGTGGGGGCTGATCTTGGG

12361
ACAAGAAGACAAATAATCCAGGCTCTCTGTCTCACACCAGCTGCCCCCCCCCTTTCTTCC
TGGCACAGTCATGTTGGAACCAGCTGCTGAGACCATTCCTCAGACCCAGAGATCATCAG
TGTACAGCTGCAGGTGCTTAGTCTCTTGCCCGCTACAGACAGTCCATCTCTGTTCTGGC
CGGGTCCACCTGGAGAGATGCTCTGAAGAAGGCCATGAGTTAGGAGGATTCAGGTGAGA
CTCCACCTCCAGTCTCACCCCAACCTCACATGCCTGATAACAGGGTCACAGAA
[A, G]
AGACGGGGAACAGAGGAGAGGGTTCCCTCGGGAGAGACACTGGCCCTGCTTCTGCTTCTA
CCTGCTCAGCTCCTTTCTTGGCCACGGTGTATGGAACAGGGAGCCATAGGCCAGCATT
GTCACIGAGAGAGCAGGCTTTGGAGGCAGAGCCCCCAGTTGGAATCCCACTCTAACCA
GCTAGGTTCCAGGTAGGCACCCACAATTACCGAGGAGAACAGTTGTGCCCTTCCCTGC
AGGGCCAGTGTGAAGAGTCCAGGAGTTAGTACACATAGAGATAGTGGCATGTGCTTTT

13187
GGCACGTGCCACACAACCTGGGTAATTTTTTTTTTTTTTTTTTTTGGATAGGGTCTCTG
TCTGTTGCCAGGCTGGTCTCAAATTCCTGGCCCTCAAACCATCTCACACCTGAGGCGCT
CAAAATATTGGGATTATAGGTGCGAGCCATCATGCTCAGCCAGAATAATACTGGTTTTT
TTGTTTTTTTTTTTGGACAGAGTCTCACTCTATTACCAGGCTCTGGAGGCCAACTCG
TGTTTTGTGATTGTTTTATTTATTTATTTATTTATTTTCGAGACAGAGCCTCTCTCTT
[T, C]
ACCTAGGCTGGAGTGCAGTGGCGCAATCTCGGCTCACTGCAACCTCCGTCTCTGGGTTT
AAGTGATTGTCTGCTCAGCCTCCTGAGTAGCTGGTGCTACAGGCGCGTGCCACCATGC
CCAGCTAATTTTTGTATTTTATAGTAGAGACAGGTTTTACTATGTTGGCCAGCTGGTTT
TAACTCCTGAACTCGGGTGATCTGCCTGCCTCGGCTCCCAAAGTGTGGGATTACAGGC
ATGGGCTCCGTGCCCGGCATGTATTTATTTAGGCAAGGTCTCTCTGTTATCCAGGC

13284
ACCATCTCACACCTGAGGCGCTCAAAATATTGGGATTATAGGTGCGAGCCATCATGCTC
AGCCAGAATAATACTGGTTTTTTTTTTGTTTTTTTTTTGAGACAGAGTCTCACTCTATTAC
CCAGGCTCTGGAGGCCAACTCGTGTGTTGTGATTGTTATTTATTTATTTATTTAT
TTCGAGACAGAGCCTCTCTTTTACCTAGGCTGGAGTGCAGTGGCGCAATCTCGGCTCA
CTGCAACCTCCGTCTCTGGGTTCAGTGATTGTCTGCCTCAGCCTCCTGAGTAGCTGG
[T, C]
GCTACAGGCGGTGCCACCATGCCAGCTAATTTTTGTATTTTATAGTAGAGACAGGTTTT
TACTATGTTGGCCAGCTGGTTTTCTAACTCCTGAACTCGGGTGATCTGCCTGCCTCGGCCT
CCCAAAGTGTGGGATTACAGGCATGGGCTCCGTGCCCGGCATGTATTTATTTAGGCA
AGGTCTCTCTGTTATCCAGGCTGAAGTGCAGTGGCACATTATAGCTCACTGCAGCCT
CAAATTATCCAAGTAACAGGGACTACAGGCATGCACCACCACCCATCTACTTTTTTTT

21341
TCAGTCACTGCAACCTCTGCCTCCCTGGCTTAAGCGATCCTCCCACCTCAGCCTCCTGA
GTACGTGTGACCATAGGCCATGGCACAAAGCCAGCTAATTTTTTTGATTTTATAGTAGA
AATGTGGTTTACCATGTTGCATAGGCTGGTCTCGAACTTCTGAACTCAAGTGATCTGCC
TGCCTTGGCCTCCCAAAGTGTGGGATTCTAGGTATGAGCCACCCTGCTCGGCCTATAAT
GGCACTTTCTATCCCATGATGAGGCTCTACTCTCATGACCTAATCATCTCCCAAAGGC
[C, G]
CTAAGGCCTCCTGATACCATCACCTTTGGGGTTAGGTTTTAAACATATACATTTGGGGG
ACACAGACATTTAGACCATAGCACCTCCATTGAAAGGAAACATTTCTGACACCTGGCTA
TCTCAAAGGGCCCTTTCACTTCCCTGCAGGCTGCATTTCCACATCACCACCAAGAGCAG
CGACACTCACTCAGAGGTTAAATAACTTGTCCAGAGTCACAGCAGTAATGAATGACAGAG
CTGGGCTTGAATCCAGGCTCCCTAGAGCCTGGATTCTGTGTAGTGAGTGAAAGCTG

21796
CATTTCCACATCACCACCAAGAGCAGCGACACTCACTCAGAGGTTAAATAACTTGTCCAG
AGTCACAGCAGTAATGAATGACAGAGCTGGGGCTTGAATCCAGGCGTCTCTAGAGCCT
GGATTCTGTGTAGTGAGTGAAAGCTGACTCCTGGGAGACTTCTGCGTGGTCTGGTCTC
TCTCCAGACTGCACTGCGCAAGTTCTCTTCTGATGGTCCCTAGGGTATTACAAAGACA

FIGURE 3, page 19 of 22

GTGGCCCTGCCTGTCAGGTGTTTTATTACCAGATGAGGTCATGGCCTCAGGAACCCGTGT
[A, G]
GGAAGCTGAGTTTACAGACTCTTTGAGCAGGCTTTAGGGAGGTTCCAGCTTCCCACCACCAA
GCCCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGTGGGGGTCTGGAAGATGAGGTTG
CGGGTGGCATATTCTGCCCAATTGCCCCCTCCTTGCTCAATCTGTTTCTGCAGGTATTG
CTGACTACAGACCCCAAGGATGGAGAAACCATTGAGCTGAGGCTGGTTAGCTGGTAGCCCC
TGAGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTGATG

22159 CCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGTGGGGGTCTGGAAGATGAGGTTGCG
GGGTGCGATATTCTGCCCAATTGCCCCCTCCTTGCTCAATCTGTTTCTGCAGGTATTGCT
GACTACAGACCCAAGGATGGAGAAACCATTGAGCTGAGGCTGGTTAGCTGGTAGCCCCTG
AGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTGATGTC
CCTGGAACAGGAACTCGCCTGACCTGCTGCCACTCCTGTGCACTTTGAGCAATGCCCC
[C, T]
TGGGATCACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGAGCAG
AGAGCCAAGCATCTTCCCTGGGAAGTCTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCA
GGTCTCCCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGGC
AAAAACGGAGTCCGAGGCCGAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTC
CTGCAAGAAGGCCCTCCTCAGCCGCGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCT

22334 CCCTGAGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTG
ATGTCCTTGGAAACAGAACTCGCCTGACCCTGCTGCCACCTCCTGTGCACTTTGAGCAAT
GCCCCCTGGGATCACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTG
GAGCAGAGAGCCAAGCATCTTCCCTGGGAAGTCTTCTGGCCAAGTCTGGCCAGCCTGGC
CCTGCAAGGTCTCCCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTC
[C, A]
TTGGCAAAAAACGGAGTCCGAGGCCGAGGTGTTGTGAAGACCACTCGTTCTGTGGTTG
GGGTCTGCAAGAAGGCCCTCCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACT
CTGCTGTTAGAGTGGCAGCTCCGAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCA
GGGCTGCTCAGTGCCCTGCTCTGACAAAATTAAGCATTGATGGCCTGTGGACCTGTAC
AGTGGCCTGGTGCCTCATACTCCTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGAAGTA

22346 TCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTGATGTCCCTGGAA
CAGGAACCTCGCCTGACCCTGCTGCCACCTCCTGTGCACTTTGAGCAATGCCCCCTGGGAT
CACCCCTGGGATCACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGAGCAGAGAGCC
AAGCATCTTCCCTGGGAAGTCTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCAGGTCTC
CCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGCCAAAAA
[C, -]
GGAGTCCGAGGCCGAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTCTGCAAG
AAGGCCCTCCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCTGTTAGAG
TGGCAGCTCCGAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCAGGGCTGCTCAGT
GCCTGCCTCTGACAAAATTAAGCATTGATGGCCTGTGGACCTGTCTACAGTGGCCTGGTG
CCTCATACTCCTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGAAGTAACCCCATCAGGG

22476 ACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGAGCAGAGAGCCAAGCATCTTC
CCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCAGGTCTCCCATGAAGGC
CACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGCCAAAAACGGAGTCCGC
AGGCCGAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTCTGCAAGAAGGCCCTCC
TCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCTGTTAGAGTGGCAGCTC
[C, T]
GAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCAGGGCTGCTCAGTGCCTGCCTCT
GACAAAATTAAGCATTGATGGCCTGTGGACCTGTCTACAGTGGCCTGGTGCCTCATACTC
CTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGAAGTAACCCCATCAGGGAGGAGTGGAG
GGTGCCTGAGCCGCATGTGGGCATTGGGGAGTGATGGGAATGCCAGCAGTGATGACGT
TGACTACTGACTGAGCACCCTACTATGACTGAGCACTCACTCGCTAGATACTATCTTG

23143 GCCGGGCGTGGTGGCTCACGCCCTGTAATCCAGCACTTTAGGAGGCCAAGGCAGGTGGAT
CACAAGGTGAGGAGTTTGAGATCAGCCTGGCCAACATGGTGAACTCCATCTTTACTAAA
AATACAAAAAATTAGCCAGGCATGGTGTTCATGCCTGCATGCCTGTAATCCAGTTACT
TGGGAAGCTGAGGCAGGAGATTGCTTGAACCCCTGGAGGCGGAGGTTGATGAGCCGAG
ATCACGCCATTGCACTCCAGCTTGGGCAAGAAGAGAAACACTCTCAAAAAAAAAAAAAA
[A, -, T]
CAGGAAACTGGTGCTCAAAAAGGAAAAGTGACTCACCAAGGTCACAGACTAGGCAGTGAT
GCTGGGGGAACCTGGCTCAGGGGACACAGACCTGGCCCTGGGGCAGCCTTGCAGCTCCTCC
ACTAAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAA
CTCAACTGGAGCCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTTCCACAAGGCA
CCATTAGGGCATCACCCAGGCCCTCGGCAGCCACGACGAGGATCCTGCCTCTCATTGG

23445 AGGAAACTGGTGCTCAAAAAGGAAAAGTGACTCACCAAGGTCACAGACTAGGCAGTGATG
CTGGGGGAACCTGGCTCAGGGGACACAGACCTGGCCCTGGGGCAGCCTTGCAGCTCCTCCA
CTAAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAAC

FIGURE 3, page 20 of 22

TCAACTGGAGCCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTTCCACAAGGCAC
CATTAGGGCATCACCCAGGCCCTCGGCAGCCACGACGAGGGATCCTGCCTCTCATTGGT
[T, C]
GGGGGCTTAGGGGCTCTGGGCTGCCCTCTTGAAGAGGGGTTACAGCCAGCGAGGCACCC
CCTATGCTGCACCCACCAAGGTAGGAAGAGGTCTGTCTCAGTGGGGCCCTCTGATG
AACAGCCCATCAGGTCTGCGTCCACATGCCTTGAAGAGATGGTGACATACTCAAAGTCC
TTGAAGCCGCATATTAAACCACCTAGAGCACCATCTTCAAACATTTAGGGTCTGAGAAGA
TAGGGGAAGTAAGCAATTTAAACATTTCTTTATATTGGGCCAGGTGCAATGGCTCACGT

23974 GGTCTGAGAAGATAGGGGAAGTAAGCAATTTAAACATTTCTTTATATTGGGCCAGGTGC
AATGGCTCACGTCTGTAATCCCAGCGCTTTGGGAGGACGAGGATCACCTGAGGTGAGGAG
TTCAGATCAGCCTGGCCAACATGGAGAAACCCATCTCTACTAAAAATACAAAATTAG
CTCAGGCGTGGTGATGTGCACCTGTAATCCTAGCTATTAGGAGGCTGAGGCACAAGAA
TGCTTGAGTCAATATTGCACCACTGCACCTCCAGCCTGGGCAACAGCGAGACTCTTGCTC
[A, -]
AAAAAAAAAAAAAGATATTTGCTGAAAAGACCCAGCCTGCCAAACTCAGGGGCAGCCAAGG
GAGGTAGTGAAATGGAAGTTGGAGCTCAGCGCTCCACACCTCCACTGCCCTCAGGCCTT
CTCTGCCTCTTTCCCATCAGTCAGTGTCTTCTGGGCATGGTCTTGGCAGAGACTTGGCCT
CCTTCCAGTTCAAGCTCCCTCTTAGATTGTGTCCACGCCACTGAGTCTTTGGGACACTG
GGTCAGATGTCTAGTCTGGCACAATTGGCAGGAATCCAAGAAACAGTGTGAGTGAGGGG

24865 ATAGCTGGGATTACAGGTGTGCACCACCATGCCAGCCTAATTTTGTATTATTAGTAGA
GATGGGCTTTTACCATGTTGTCCAGGCTGGTCATGAACCTCCTGACCTCAAGTGATCCACC
CGCTTTGGCCTCCCAAAGTGTGGGATTACAAGCATGAGCCACAGTGCCTGGCCTGACCC
TGCTCTTTTGAAGACCATTCCCCCAAATCTGTGCACCTGTGTGCCTTTCTCTCTCTG
CCTCCTCTCAGCTCTGCCCCGCTCTCCTCCCTTCTCCTCTGGCAAATCCCACTCATCTCT
[G, T]
GAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGGATGAAG
GACGTGGCCACGGAGTTTGTGTTTGTGTTTGTGAGATGGAGTTTGTCTCATGTTGCCC
AGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGTTCAAGC
GGTTCTCCTGCTTAGCCTCCCAAGTGTGGGATTACTGGCATGAACCACCACACCTGG
CTAATTTTGTGTTTTTGTAGTAGAGATGGGGTTTCTTCATGTTGGTCAGGCTGGTCTCGAAC

25040 GACCCTGCTCTTTTGAAGACCATTCCCCCAAATCTGTGCACCTGTGTGCCTTTCTTCT
CTCTGCCTCTCTCTCAGCTCTGCCCCGCTCTCCTCCCTTCTCCTCTGGCAAATCCCACTCA
TCTCTTGAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGG
ATGAAGGACCTGGCCACCGAGTTTGTGTTTGTGTTTGTGAGATGGAGTTTGTCTCATG
TTGCCCAGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGT
[T, C]
CAAGCGGTTCTCCTGCCTTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCACCACA
CCTGGCTAATTTTGTGTTTTTAGTAGAGATGGGGTTTCTTCATGTTGGTCAGGCTGGTCT
CGAATCCCAACCTCAGGTGATCTGCCTGCCTCGGCCTCCCAAAGTACTGGGATTACAGG
GTTGAGCCACTGTGCCTGGCCAGGCCACCGAGTTTAAAGAGGCTTCTGTGGCAGTGG
CATCCAGACGGAGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGGAGTG

25046 GCTCTTTTGAAGACCATTCCCCCAAATCTGTGCACCTGTGTGCCTTTCTTCTCTGTC
CTCCTCTCAGCTCTGCCCCGCTCTCCTCCCTTCTCCTCTGGCAAATCCCACTCATCTCTT
GAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGGATGAAG
GACGTGGCCACGGAGTTTGTGTTTGTGTTTGTGAGATGGAGTTTGTCTCATGTTGCCC
AGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGTTCAAGC
[G, A]
GTTCTCTGCCTTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCACCACACCTGGC
TAATTTTGTGTTTTTAGTAGAGATGGGGTTTCTTCATGTTGGTCAGGCTGGTCTCGAACT
CCCAACCTCAGGTGATCTGCCTGCCTCGGCCTCCCAAAGTACTGGGATTACAGGTTGAG
CCACTGTGCCTGGCCAGGCCACCGAGTTTAAAGAGGCTTCTGTGGCAGTGGCATCCA
GACGGAGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGGAGTGTGAGTT

25193 ATCATGCTGCTTTCTCCTGTGGGAGGGATGAAGGACGTGGCCACGGAGTTTGTGTT
TTGTTTTGAGATGGAGTTTGTCTCATGTTGCCAGGCTGGGGTACAATGGTACGATCTCA
GCTCACTGCAACCTTACGTCCCGGGTTCAAGCGGTTCTCCTGCCTTAGCCTCCCAAGTA
GCTGGGATTACTGGCATGAACCACCACACCTGGCTAATTTTGTGTTTTTAGTAGAGATGG
GGTTTCTTCATGTTGGTCAGGCTGGTCTCGAACTCCCAACCTCAGGTGATCTGCCTGCCT
[C, T]
GGCCTCCCAAAGTACTGGGATTACAGGTTGAGCCACTGTGCCTGGCCAGGCCACGGGA
GTTTTAAGAGGCTTCTGTGGCAGTGGCATCCAGACGGAGTGCAGAACTCAAAGTTGAA
GGCCAGAAGCTCAGGGAAGGGGGAGTGTGAGTTGAGGAGTCTCTTGGCTGCCAGGGCCAG
AAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAGAGCATGTAGAATCAGAGAGGAG
GCTGAGCCATGCAGCCCGAGAAGAGGGGAATGCCACTGAGCCACAGAGACCCAGTGCCA

25529 AGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGGAGTGTGAGTTGAGGA
GTCTCTTGGCTGCCAGGGGCCAGAAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAG

FIGURE 3, page 21 of 22

AGCATGTAGAATCAGAGAGGAGGCTGAGCCATGCAGCCCCGAGAAGAGGGGAATGCCACT
GAGCCACAGAGACCCAGTGCCACTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCC
[T, G]
GCCTGTCTCTGTATGCAAGGCTTCACCCCTCTCTCGTTGTACATTGTACACATTCTAGGTGA
CACCAGCAGCTTCTGATTCTCATCTCCCATAAATCAGCCCCCAGAGAGGGGACAACCTG
CTGAGCTGATAACATAATAGATGCCCTTTCTTGAGGCCATGGTCATGGTCAGCGTGGA
GAGGATGAAGCCTGAGCAGGCAGGATCGGGGGTCTAGAGGGGAAGGAGGTGGAAGTT

25614 GGCAGAAAGCTCAGGGAAGGGGGAGTGTGAGTTGAGGAGTCTCTTGGCTGCCAGGGCCAG
AAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAGAGCATGTAGAATCAGAGAGGAG
GCTGAGCCATGCAGCCCCGAGAAGAGGGGAATGCCACTGAGCCACAGAGACCCAGTGCCA
CTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCCGGCCTGTCTCTGTATGCAAGGCTTCA
CCCTCTCTCGTTGTACATTGTACACATTCTAGGTGACACCAGCAGCTTCTGATTCTCATC
[C, T]
CCCATAACTCAGCCCCCAGAGAGGGGACAACCTGCTGAGCTGATAACATAATAGATGCC
CCTTTCTTGAGGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGAGCAGGCAGGA
TCGGGGTCTAGAGGGGAAGGAGGTGGAAGTTGAGATCACAGACCTGTGGTCAGGTGGCC
TGGGAAGGGTTTGACAGGTGTCGGCCCAAAGAGCTTGAAGGGATTTTGCTGCTGTGGGT
GAGCACTGCCTCTCCCTTAGGGACAACAGCCACCTCTTCTCTCCCATTTGCTTTTCCC

25785 CCAGTGCCACTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCCGGCCTGTCTCTGTATG
CAGGCTTCACCCCTCTCTCGTTGTACATTGTACACATTCTAGGTGACACCAGCAGCTTCTG
ATTCTCATCTCCCATAACTCAGCCCCCAGAGAGGGGACAACCTGCTGAGCTGATAACAT
AATAGATGCCCCCTTTCTTGAGGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGA
GCAGGCAGGATCGGGGGTCTAGAGGGGAAGGAGGTGGAAGTTGAGATCACAGACCTGTGG
[C, T]
CAGGTGGCCTGGGAAGGGTTTGACGAGTGTGGCCCAAAGAGCTTGAAGGGATTTTGCT
GCTGTGGGTGAGCACTGCCTCTCCCTTAGGGACAACAGCCACCTCTTCTCTCCCATTT
GCCTTTCCCTTCTGTAGATATGAAACACAGGCCTCCTTGTGAGCCCCCTACTTAACCTCC
GTGATGGGGAAGCGGCCGGAGAAAGGGAGTTCTGGCAGCTTCTCCGAGACCCCAACACC
CCACTGTTGCAAGGTGAGTCATGGCCTGACACTCTGGATGTGTCCCTACCCCAAGCTTA

26266 GTGATGGGGAAGCGGCCGGAGAAAGGGAGTTCTGGCAGCTTCTCCGAGACCCCAACACC
CCACTGTTGCAAGGTGAGTCATGGCCTGACACTCTGGATGTGTCCCTACCCCAAGCTTA
CTCAGCCAAGAGGCTTCATCAACTCAGCCAGCTTTCCCTAGCACCCCTCCTGGGCCACAC
CTTCACAAAATCACTGATGCTCAAAGTTGGATATAATATATTGAAGTGAAGCCTTAGCAT
TTTTATGCAAGTTACTGTGGAATTTCTAGGAAACCAGACAGATTACAAAAA
[A, -]
CTAGAAGAAAATTAACATCACCTAGGATATACTACCTAGGAATAACGTCTTTTATTTGA
GATGGAGTTTCGCTCTTGTGGCCAGGCTGGAGTGCAGCGGTATGATCTCGGCTCGCTGC
AACCTCCGCCTCCTGGGTTTATGTGATTCTTCCACCTCGGCCTTCTAGAGCCCAAGTGG
TCGCTGCTGCTCTGCTCCCAAAGTTCTGGGATTACAGGCATGAGCCACCGCACCCAGCC
AAAATTACTTAACCTTTCTTCTAGATACTTTTAAAAATATGGCAGTAAGTTTTTCATAA